



**Community Forum Deliberative Methods
Demonstration: Evaluating Effectiveness and Eliciting
Public Views on Use of Evidence**

Final Report



Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Community Forum Deliberative Methods Demonstration: Evaluating Effectiveness and Eliciting Public Views on Use of Evidence

Final Report

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2010-00005

Prepared by:

American Institutes for Research
Washington, DC

AHRQ Publication No. 14(15)-EHC007-EF

November 2014

The Agency for Healthcare Research and Quality's (AHRQ) Effective Health Care Program conducts and supports research focused on the outcomes, effectiveness, comparative clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services. More information on the Effective Health Care Program and electronic copies of this report can be found at www.effectivehealthcare.ahrq.gov.

The Community Forum was initially funded under the American Recovery and Reinvestment Act (Public Law 111-5) to expand public and stakeholder engagement in AHRQ-supported comparative effectiveness research programs. This final report was developed with ARRA funding, by the American Institutes for Research, under AHRQ Contract No. 290-2010-00005.

The AHRQ Task Order Officer for this project was Joanna Siegel, Sc.D. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ or the U.S. Department of Health and Human Services. Therefore, no statement in this report should be construed as an official position of AHRQ or the U.S. Department of Health and Human Services.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

People using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

James Fishkin at the Center for Deliberative Democracy, Stanford University, was the original developer of “Deliberative Polling[®],” one of the methods included in this study. No other interest or affiliation is noted among other members of the study investigator team.

Suggested Citation:

Carman KL, Maurer M, Mallery C, Wang G, Garfinkel S, Richmond J, Gilmore D, Windham A, Yang M, Mangrum R, Ginsburg M, Sofaer S, Fernandez J, Gold M, Pathak-Sen E, Davies T, Siu A, Fishkin J, Rosenberg M, Fratto A. Community Forum Deliberative Methods Demonstration: Evaluating Effectiveness and Eliciting Public Views on Use of Evidence. (Prepared by the American Institutes for Research under Contract No. 290-2010-00005.) AHRQ Publication No. 14(15)-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2014. www.effectivehealthcare.ahrq.gov.

Acknowledgments

The American Institutes for Research (AIR) would like to acknowledge the U.S. Department of Health and Human Services' Agency for Healthcare Research and Quality for sponsoring this project. We also thank our Project Officer, Dr. Joanna Siegel, for her leadership and guidance.

We extend gratitude to our multidisciplinary team of organizations and individuals for their expertise, input, and commitment to this 3-year initiative to improve and expand public and stakeholder engagement in the Agency's Effective Health Care (EHC) Program. Below, we list individuals who contributed to the design, implementation, and/or evaluation of the Community Forum Deliberative Methods Demonstration.

American Institutes for Research (in alphabetical order)

Kristin Carman, Ph.D., Project Director

Amar Beeharry
Matthew Borda, M.A.
Johannes Bos, Ph.D.
Graciela Castillo, M.P.H.
Michael Cohen, Ph.D.
Sandy Eyster, Ph.D.
Jessica Fernandez
Kirsten Firminger, Ph.D.
Allison Fratto
Samuel Mintzer Fuchs
Steven Garfinkel, Ph.D.
Dierdre Gilmore, M.A.
Susan Heil, Ph.D.
Jessica Heeringa, M.P.H.
Kyosin Kang
Chelsea Kelleher
Laurel Koester, M.P.H.
Elena Liroff
Coretta Mallery, Ph.D.
Rikki Mangrum, M.L.S.

Diane Martinez, M.P.H.
Regin Mathew
Maureen Maurer, M.P.H.
Marilyn Moon, Ph.D.
Peter Morrison
HarmoniJoie Noel, Ph.D.
Alex Ortiz
Ushma Patel, M.S.P.H.
Jennifer Richmond
Simone Robers, M.A.
Mark Rosenberg
Herschel Sanders
Karthik Shyam, M.P.P.
Adair Smith
Andrew (Kip) Thomson
Grace Wang, Ph.D., M.P.H.
Channet Williams
Amy Windham, Ph.D.
Thomas Workman, Ph.D.
Manshu Yang, Ph.D.
Jiani Yu

Center for Healthcare Decisions

Fern Alexander, Ph.D.
Tamar Foster, M.P.P.A.
Marge Ginsburg, M.P.H.
Kathy Glasmire

Center for Deliberative Democracy, Stanford University

James Fishkin, Ph.D.
Alice Siu Chao, Ph.D.
Kathleen Giles, M.A.

Center for the Study of Language and Information, Stanford University

Noam Ben-Avi
Todd Davies, Ph.D.
Michael Huang
Mike Mintz, M.S.
Kenneth Taylor, Ph.D.
Jimmy Tobin

Consultants

Marthe Gold, M.D., M.P.H., City College of New York
Ela Pathak-Sen, M.B.A., Commotion
Shoshanna Sofaer, Dr.P.H., Baruch College, City University of New York

Technical Expert Panel

Julia Abelson, Ph.D., McMaster University
Stirling Bryan, Ph.D., University of British Columbia
Tim Carey, M.D., M.P.H., University of North Carolina–Chapel Hill
Scott Y. Kim, M.D., Ph.D., University of Michigan
Peter Muhlberger, Ph.D., Texas Tech University
David Ropeik, M.A., Harvard University
Mark E. Warren, Ph.D., University of British Columbia

Expert Presenters for the Deliberative Demonstration

Mark Holmes, Ph.D., University of North Carolina–Chapel Hill
David Hyman, M.D., J.D., University of Illinois
Anthony Iton, M.D., J.D., M.P.H., The California Endowment
Lawrence Kleinman, M.D., M.P.H., Harvard
Christine Mitchell, R.N., M.S., MTS, FAAN, Children's Hospital Boston
John Morton, M.D., M.P.H., M.H.A., FACS, FASMBS, Stanford University
Ira Moscovice, Ph.D., University of Minnesota
Rita Redberg, FACC, M.D., M.Sc., University of California, San Francisco
Kyu Rhee, M.D., M.P.P., FAAP, FACP, IBM
James Sabin, M.D., Harvard University Medical School
Lisa Schwartz, M.D., M.S., Dartmouth
Mark Tonelli, M.D., M.A., University of Washington
Glennah Trochet, M.D., Sacramento County
Steven Woloshin, M.D., M.S., Dartmouth

Contents

Executive Summary	ES-1
Section I: Implementation and Research Design	1
Chapter 1. Introduction	2
Background on Public Deliberation.....	2
Aims and Study Design	4
Organization of This Report	5
Chapter 2. Design and Implementation of the Deliberative Methods Demonstration	6
Public Deliberation: Defining Principles and Core Elements	6
Designing Four Deliberative Methods.....	9
Developing the Deliberative Topic	12
Educational Materials and Experts	13
Creating the Community Forum Web Site	18
Facilitating the Deliberative Sessions.....	19
Pilot Testing the Methods	20
Implementation	21
Chapter 3. Research Methods.....	23
Aim 1: Inform AHRQ Programs on Public Views About the Use of Evidence in Health Care Decisionmaking.....	23
Aim 2: Expand the Evidence Base on Public Deliberation	29
Strengths and Challenges of the Community Forum Deliberative Methods Demonstration ...	50
Section II: Research Findings	54
Chapter 4. Answering the Deliberative Question: The Public's Perspective	55
Background	55
Findings	57
Discussion	77
Chapter 5. Research Question 1: Comparing Effectiveness of Public Deliberation With Receiving Educational Materials.....	81
Methods	82
Results.....	85
Summary and Discussion.....	92
Chapter 6. Research Question 2: Participant Deliberation Quality and Experience.....	96
Methods	96
Results.....	97
Discussion	98
Chapter 7. Research Question 3: Comparing Effectiveness of Specific Deliberative Methods.....	99
Methods	100
Results.....	104
Discussion	108
Chapter 8. Research Question 4: Examining Effectiveness of Public Deliberation by Participants' Personal Characteristics	111
Methods	112
Results for Research Questions on Personal Characteristics and Concordance (RQ4.1 and RQ4.2).....	117

Discussion	134
Conclusion	138
Chapter 9. Research Question 5: Group Behavior—Consensus, Polarization and Domination	139
Data Sources and Methods	140
Consensus	142
Polarization	143
Domination	147
Discussion	150
Chapter 10. Evaluation of Costs	152
Costs Included and Excluded—and Why	152
Analysis of Costs	155
Results	155
Discussion	156
Section III: Implications.....	158
Chapter 11. Discussion and Implications	159
Public Input on the Use of Evidence in Decisionmaking	159
Expanding the Evidence Base on Public Deliberation	160
Continuing the Learning: Areas for Future Research	163
Conclusion	165
References	166

Exhibits

Exhibit 1-1. Core elements of public deliberation	3
Exhibit 2-1. Approaching the elements that promote deliberation quality and effectiveness	9
Exhibit 2-2. How the methods varied	10
Exhibit 2-3. Commonalities and differences among the deliberative methods	12
Exhibit 2-4. Participant educational materials for each deliberative method	13
Exhibit 2-5. Groups by month, method, and location.....	22
Exhibit 3-1. What defines the quality of discourse?	28
Exhibit 3-2. Participant selection of deliberative methods during screening: frequency of different combinations	35
Exhibit 3-3. Number of participants targeted, number recruited, and number who showed	38
Exhibit 3-4. Participant show rate by method, location, and demographic characteristics	39
Exhibit 3-5. Who responded to the Knowledge and Attitudes Pre and Post Surveys?.....	41
Exhibit 3-6. Who responded to the Deliberation Quality and Experience Survey?	42
Exhibit 3-7a. Final outcome measures for knowledge of health care issues presented in educational materials	45
Exhibit 3-7b. Final outcome measures for attitudes toward using medical evidence in health care decisionmaking.....	46
Exhibit 3-7c. Final outcome measures for considerations of costs in health care decisionmaking	46
Exhibit 3-7d. Final outcome measures for attitude toward mechanisms to support use of high-volume hospitals	47
Exhibit 3-7e. Final outcome measures for attitude toward the use of medical evidence to limit antibiotic use (restricted to Community Deliberation and Citizens' Panel methods)	47
Exhibit 3-7f. Final outcome measures for attitudes toward the role of society in prevention (restricted to Citizens' Panel method).....	47
Exhibit 3-7g. Final outcome measures for deliberation quality and experience: quality of communication and discourse.....	48
Exhibit 3-7h. Final outcome measures for deliberation quality and experience: implementation process	48
Exhibit 3-7i. Final outcome measures for deliberation quality and experience: participant ratings of the impact of deliberation	49
Exhibit 4-1. Demographic characteristics of participants in deliberation.....	56
Exhibit 4-2. Summary of case study examples	57
Exhibit 4-3. Participant perceptions of boundaries	61
Exhibit 4-4. How do participants perceive boundary-setters?	63
Exhibit 4-5. How participants perceived different levels of evidence of medical effectiveness.....	66
Exhibit 4-6. How participants perceived different types of evidence of harm	68
Exhibit 4-7. Using evidence as a basis for boundary setting approaches	71
Exhibit 4-8. What kinds of values were exhibited during deliberation?.....	73
Exhibit 5-1a. Knowledge outcome measures.....	83
Exhibit 5-1b. Attitude outcome measures common to all deliberative methods	83
Exhibit 5-1c. Attitude outcome measures evaluated only for Community Deliberation and Citizens' Panel	83
Exhibit 5-1d. Attitude outcome measures evaluated only for Citizens' Panel	84

Exhibit 5-2a. Effect of deliberation on knowledge: pooled deliberative groups versus reading materials only control	86
Exhibit 5-2b. Effect of deliberation on attitudes toward using medical evidence in health care decisionmaking: pooled deliberative groups versus reading materials only control	87
Exhibit 5-2c. Effect of deliberation on attitudes toward mechanisms to support use of high-volume hospitals: pooled deliberative groups versus reading materials only control	87
Exhibit 5-3a. Effect of deliberative methods on knowledge: each deliberative method versus control	89
Exhibit 5-3b. Effect of deliberative methods on attitudes toward using medical evidence in health care decisionmaking: each deliberative method versus control	90
Exhibit 5-3c. Effect of deliberative methods on attitudes toward mechanisms to support use of high-volume hospitals: each deliberative method versus control.....	91
Exhibit 5-3d. Effect of deliberative methods on attitudes toward the use of medical evidence to limit antibiotic use: each deliberative method versus control.....	92
Exhibit 5-3e. Effect of deliberative methods on attitudes toward the role of society in prevention: each deliberative method versus control.....	92
Exhibit 6-1. Deliberation quality and experience outcome measures.....	96
Exhibit 6-2. Deliberation quality and experience outcome measure mean scores for each method.....	98
Exhibit 7-1. Sample sizes for RQ3 analyses	101
Exhibit 7-2. Knowledge, attitude, and deliberation quality and experience outcome measures .	102
Exhibit 7-3. Effect of deliberation intensity on knowledge, attitudes, and deliberation quality and experience (CP versus BCD; CP as reference)	106
Exhibit 7-4. Effect of deliberation mode on knowledge, attitudes, and deliberation quality and experience (CD versus ODP; ODP as reference).....	107
Exhibit 8-1a. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Chicago, IL	113
Exhibit 8-1b. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Durham, NC.....	113
Exhibit 8-1c. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Silver Spring, MD.....	113
Exhibit 8-1d. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Sacramento, CA	113
Exhibit 8-2. Attitude and experiences outcomes	115
Exhibit 8-3. Numbers of participants completing the Knowledge and Attitudes and Deliberation Quality and Experience Surveys, by demographic characteristics	117
Exhibit 8-4. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by race.....	119
Exhibit 8-5. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by ethnicity	123
Exhibit 8-6. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by age.....	127
Exhibit 8-7. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by educational attainment.....	131
Exhibit 8-8. Summary of findings showing strength of association between knowledge, attitudes, and experience outcomes and demographic factors.	135

Exhibit 8-9. Summary of findings on strength of association between concordance and knowledge, attitudes, and experience outcomes.	136
Exhibit 9-1a. Measures used for groups that were not used for individual-level analyses: Government.....	141
Exhibit 9-1b. Measures used for groups that were not used for individual-level analyses: Factor 2a – Doctors should consider evidence over preferences when making treatment decisions.....	141
Exhibit 9-1c. Measures used for groups that were not used for individual-level analyses: Factor 2b – Patients should consider evidence over preferences when making treatment decisions.....	141
Exhibit 9-1d. Measures used for groups that were not used for individual-level analyses: individual indexes	141
Exhibit 9-2. Number and percent of groups moving toward consensus by method and measure	142
Exhibit 9-3. Mean change in standard deviation by deliberative method and measure	143
Exhibit 9-4. Illustration of polarization	144
Exhibit 9-5. Percentage of group-by-measure opportunities moving toward the extremes, by deliberative method.....	146
Exhibit 9-6. Illustration of domination	148
Exhibit 9-7. Definition of social privilege items	148
Exhibit 9-8. Percentage of groups that moved at T2 toward the T1 position of privileged members.....	150
Exhibit 10-1. Number of participants and total hours of deliberation, by method	153
Exhibit 10-2. What costs were associated with holding deliberative sessions?	153
Exhibit 10-3. Costs per deliberative session: base estimate, low- and high-range estimates, by method (2012 U.S. dollars).....	155
Exhibit 11-1. Overall assessment of outcomes and investment, by method.....	162

Appendixes

Appendix A. Technical Expert Panel

Appendix B. *Preparing for the Community Forum* Booklet

Appendix C. Case Studies

Appendix D. Screenshots of the Deme Web Site

Appendix E. Final Qualitative Code List

Appendix F. Telephone Screening Form for Recruiting Participants

Appendix G. Informed Consent Form for In-Person Discussion Groups

Appendix H. Technical Description of Randomization and Weighting Procedures

Appendix I. Knowledge and Attitudes Survey

Appendix J. Deliberation Quality and Experience Survey

Appendix K. Summary of Responses to Case Studies

Executive Summary

Introduction

The Agency for Healthcare Research and Quality (AHRQ) Community Forum, initiated under the American Recovery and Reinvestment Act (ARRA), has as its goal to improve and expand public and stakeholder engagement in AHRQ's Effective Health Care Program. A primary area of focus for the Community Forum is to advance methods for obtaining input from the general public.

This report describes the results of the Deliberative Methods Demonstration, a randomized controlled trial comparing deliberative methods with one another and with a control intervention. The primary aims of the Demonstration were to:

- **Inform AHRQ research programs on public views regarding the use of research evidence in health care decisionmaking** by obtaining informed public input on questions regarding appropriate and acceptable ways to use evidence that are central to the mission of AHRQ's research programs.
- **Expand the evidence base on public deliberation** by evaluating whether public deliberation is an effective and useful way to obtain informed public input for U.S. health care research, as well as identifying a feasible set of choices among deliberative methods.

What Is Public Deliberation and How Has Its Effectiveness Been Evaluated?

Public deliberation methods provide opportunities to obtain informed perspectives on complex topics that are value laden and that lack simple technical solutions. On such issues, public input on what underlying values should be considered, potential tradeoffs in values, and potential solutions and their likely uptake or resistance are important considerations in developing programs or policies.

Deliberative methods are a distinct approach to obtaining public input. In public deliberation members of the public are convened to obtain input about—and meaningful insights into—how people think about a topic when they are informed. Thus, information obtained through public deliberation differs from that collected through surveys or focus groups, which generally obtain more top-of-mind—that is, initial and more intuitive—responses and reactions. In deliberative sessions, participants receive information that is intentionally neutral and respectful of the full range of underlying values, experiences, and possible perspectives. They are encouraged to discuss, learn from others, and examine and refine their own views.

Although considerable theoretical and case-study literature endorses the value of public deliberation, little empirical research has been conducted about its effectiveness.¹ In the research that has been done, effectiveness has been defined as:

- **The quality of deliberative experience or discourse.** Using participant self-reports, researchers' observations, or reviews of session transcripts, these measures typically assess levels of equal participation, active participation, opportunity for adequate discussion, respect for the opinions of others, and awareness of different perspectives.

- **Changes in participants' knowledge or attitudes about the deliberative topic.** A core goal of deliberative methods is informed input, and a core assumption is that information and discussion may alter the views of participants. Thus, typically using pre and post surveys, these measures assess the effect of the deliberation on the participants' knowledge, attitudes, perspectives, values, beliefs, opinions, or policy preferences on the deliberative topics.
- **Changes in participants' empathy and concern for issues affecting the community at large.** Using pre and post surveys, a number of studies have assessed the effect of deliberation on civic engagement and capacity, engagement in the political process, sense of self-efficacy, sense of empowerment, political efficacy and solidarity, and anticipated post-meeting activity related to deliberation issues.
- **Impact on decisions by sponsoring agency.** Ultimately, deliberation obtains information that can influence decisions. Measurement constructs include the effect of public input on specific laws, policies, or practices and on decisionmakers' intentions to act on the results of deliberation. These constructs are usually assessed through case studies or surveys of decisionmakers who may use the findings from the deliberation.

Few well-designed comparative studies of deliberative methods or their alternatives have been conducted.

Deliberative Methods Demonstration Description

Between August and November 2012, we conducted a five-arm randomized controlled trial to examine the effectiveness of public deliberation and to compare alternative approaches. Participants were assigned to one of four deliberative methods or a control group. The project convened 76 groups in four locations: Chicago, IL; Sacramento, CA; Silver Spring, MD; and Durham, NC. We selected locations that made it easier to recruit a diverse sample in terms of racial, ethnic, and sociodemographic background, with specific attention to ensuring inclusion of members of three AHRQ priority populations: Hispanics, African-American women, and the elderly.

Deliberative Topic

Across all methods, the Deliberative Methods Demonstration elicited public input on the use of research evidence in health care decisionmaking. We posed the following deliberative question to all participants:

Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?

This question was appropriate for deliberation for several reasons. First, the use of evidence in decisionmaking relates directly to AHRQ's support of research that helps people make more informed decisions and improves the quality of health care services. As such, public input on this question had the opportunity to make valuable contributions to the AHRQ program. Second, the question required participants not only to understand how evidence is generated and used, but also to discuss difficult tradeoffs concerning the impact on individuals and communities when evidence is or is not applied in medical decisions. Finally, responses to the question would elicit

the public's values around whether patients and physicians have a social responsibility to make evidence-based health care decisions.

Prior to their participation, all participants received the *Preparing for the Community Forum* booklet, which described the overall purpose of the project and what to expect (Appendix B). It also gave definitions and facts on medical research and medical evidence, quality health care, and comparative effectiveness research. Information on rising health care costs and who pays for health care was included to provide context for the discussions. We did not provide information on rules, guidelines, or any other types of boundaries in health care; rather, we allowed interpretations and discussions of boundaries to arise spontaneously.

To help participants grapple with a complex topic and a fairly abstract question, we developed specific case studies to provide context for each deliberation (Appendix C). These were:

- *Comparing Hospital Quality*
- *Upper Respiratory Infection in Children: Antibiotics Versus Symptom Treatment*
- *Obesity Management: Comparing Treatment and Prevention*
- *Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicine*
- *Comparing Approaches To Preventing Illness: A Fictional Case*

All methods used the case study on comparing hospital quality, and two methods used additional case studies.

Deliberative Methods

We selected four distinct types of deliberative methods that have been used in prior public deliberations and reflect important differences in implementation: number of participants, session length, mode of interaction, and use of content experts. We refined each type of deliberative method to ensure that all methods included necessary components of successful deliberation identified in our literature review, while retaining the methods' core distinctiveness.

Brief Citizens' Deliberation (BCD): In this method, 12 participants met in person once for 2 hours. A single facilitator and a single note-taker supported these groups. Facilitation was active, designed to encourage attention to the tensions among social values, ethical principles, and the individual versus societal perspective. Participants discussed the hospital quality case study. No expert presentations were included in this method. We held 24 BCD groups, 6 at each location.

Community Deliberation (CD): This method involved two deliberative sessions, each 2.5 hours long, 1 week apart, for each group of about 12 participants. In the first week, participants discussed the upper respiratory infection (URI) case study. During the week between in-person sessions, participants interacted through an online discussion board. In the online setting, two experts provided statements regarding the URI case study, answered participants' questions, and asked questions of their own. At the second in-person session, participants completed discussion of the URI case study and went on to discuss the hospital quality case study. A single facilitator and a single note-taker supported these groups. During the in-person sessions, facilitation was active, as described for BCD above. We held 24 CD groups, 6 at each location.

Online Deliberative Polling® (ODP): In this method, each group convened online four times, once per week over a period of 4 weeks, using one case study (hospital quality). Each meeting was a 1.25-hour online session, during which about 12 participants engaged in discussions via a

dedicated Web site and Internet-based audio conferencing. Student facilitators with no prior experience in facilitation or health care moderated these groups; they were trained to intervene as little as possible during discussions, while still attempting to ensure consideration of the competing arguments in the reading materials. This facilitation style was put in place in order to maintain the neutrality of the moderator. During the first two sessions, participants began exploring issues about hospital quality. Following discussion, the groups had the opportunity to generate questions to be addressed offline by a panel of three experts. The panelists' responses were played back to participants during the third session and served as a basis for further conversation in the final session. We held 24 ODP groups.

Citizens' Panel (CP): CP involved 2.5 days of deliberation. There were 24 to 30 participants in each group. All five case studies were used. Seven experts were linked to the group through Skype® at key points during the session to provide additional information and different points of view on the case studies and issues related to the deliberative question. A clinical expert, who was also a member of the research team, presented on comparative effectiveness research and addressed questions from participants. Three facilitators and a note-taker supported these groups. This method permitted the use of smaller breakout groups moderated by a facilitator, as well as an open space in which participants could interact without facilitation. Facilitation in this method was active, as described for BCD and CD above. We held four CP groups, one at each location.

Reading Materials Only Control Group: Participants assigned to the control intervention received educational materials via an email link. Materials included the same background booklet provided to the deliberative groups, *Preparing for the Community Forum*, as well as three of the case studies: hospital quality, URI, and obesity management. We chose three of the five case studies to present to the control (a midway point between the other methods, which received between one and five case studies). Participants did not convene in groups to deliberate. We estimated an hour of reading time.

Study Sample

Of the 1,774 participants recruited from the four locations, 961 participants took part in a deliberative method and 377 participants were a part of the reading materials only control (an overall show rate of 75 percent). The study sample was diverse and reflected each location's population in terms of sex, age, race, and ethnicity based on U.S. Census Bureau estimates, but had a larger percentage of people with at least some college education.

Findings

Public Views About Use of Evidence in Health Care Decisionmaking

To address our first aim, we conducted a thematic qualitative analysis of transcripts from the 76 deliberative groups to summarize how participants responded to the overall deliberative question. The research questions for the thematic analysis focused on three main topics related to the overarching deliberative question: (1) circumstances participants specify for restricting decisionmaking, (2) situations affecting how participants perceive those circumstances, and (3) the social values exhibited during deliberation.

When asked the overarching question, participants first focused on the concept of boundaries. Many of them initially interpreted boundaries as compulsory rules that limited choices and

allowed no exceptions, and most reacted negatively. Participants also questioned what was meant by “society.” They initially defined society as the government or a health insurance company—perceiving both types of organizations as enforcers of boundaries in health care. Participants rarely discussed the concept of evidence or questioned what the terms “evidence” or “effectiveness” meant when initially responding to the question.

Over the course of the deliberative sessions, participants expressed and debated additional viewpoints. Discussions elicited other interpretations of boundaries, including education or mandates for education, guidelines, accountability mechanisms, and penalties or incentives. Similarly, over the course of deliberation, participants discussed the relative importance of different types of evidence and the role evidence plays in decisionmaking.

Below, we summarize the main themes and values that emerged from the public’s response.

The public’s core values of individual freedom and personal choice were tempered in varying degrees by concern for the greater good or perceptions of fairness.

- The value of individual freedom emerged from participants’ consistent focus on the primacy of personal choice and negative reactions to any boundaries on decisionmaking that restrict rather than support choice. Also, participants often explicitly stated that individual freedom of choice was a core value.
- Concern for the welfare of the community at large arose when discussing evidence that unchecked individual freedom might have consequences that would harm others physically or financially. Since protecting the common good usually entailed some constraints on individual freedom, the conflict between these two values often resulted in discussion about tradeoffs. Reducing individual freedom for the good of the community was not done lightly. Concern for the greater good surfaced most clearly when discussing how blocking inappropriate use of antibiotics could prevent the development of antibiotic-resistant superbugs such as methicillin-resistant *Staphylococcus aureus* (MRSA) or how limiting patients’ choice of hospital to favor a lower volume community facility could enhance a local community’s economic well-being.
- Discussions of health care costs often elicited the value of fairness. Participants viewed fairness from a number of perspectives, including what is just in allocating shared resources and what are reasonable restrictions on patients when they are not the primary payer.

Evidence is an important component of high-quality care. Yet, given the perceived limits of applying population-based evidence to individuals, other factors often have more weight in decisions.

- In general, participants viewed evidence positively and stated that they valued it highly in making their own informed health care decisions. Participants often discussed evidence using terms such as “success rates,” “clinical results,” or “test results.” Other comments indicated that participants equated evidence with experience—the doctor’s accumulated experience and clinical judgment, personal lived experiences, or common sense. Participants’ comments indicated that knowing about unequivocal evidence and uncertain evidence is important when making an informed choice.

- When setting boundaries on decisionmaking, compelling evidence of effectiveness was necessary for encouraging better quality care, but not sufficient for constraining choice or the autonomous decisions of patients and physicians. Yet, if evidence clearly showed a treatment to be ineffective, participants were generally comfortable with setting some restrictions. In comparison, participants could not justify limiting care when the research results are mixed or the evidence itself is unclear.
- Two beliefs emerged that can act to diminish support for the role of evidence in decisionmaking. First was the view that evidence of what works for most people may not apply to each patient, as “everyone is different.” Many comments reflected participants’ perspective that evidence could be discounted if it was seen as “not applicable to me” or not applicable to the unique circumstances of specific patients in specific situations. Second, participants viewed evidence as imperfect: changing over time, often based on studies excluding specific age or ethnic subpopulations, and lacking clarity.
- Other considerations also competed with using evidence in making health decisions. Patients’ personal preferences or doctors’ clinical judgment could supersede evidence. Other features of health care—such as being treated with respect by providers, personal convenience, or concern about out-of-pocket cost—were also instrumental in determining participants’ views. Often, these other factors became more important when participants did not see the relevance of the evidence to the situation.

Evidence of physical or economic harm to individuals or the community led to increased acceptance of some limits on decisionmaking.

- When presented with the deliberative question, many participants’ initial responses showed that they perceived boundaries as compulsory rules and regulations that disallowed exceptions, interfered with the doctor-patient relationship, and limited choice. Participants expressed concerns that boundaries create logistical and practical challenges. Participants also described boundary-setting as a slippery slope, making it easier for future, inappropriate limit-setting.
- Although many participants focused on how to preserve choice and enhance the doctor-patient relationship, the majority of participants eventually concluded that some boundaries would be important or necessary to address problems in the health care system. Descriptions of harm included physical harm (e.g., pain, increased risk of future injury or illness, or death), emotional or psychological harm (e.g., anxiety about outcomes patients can expect), and economic harm (e.g., loss of community jobs, high out-of-pocket expenses for health care). Often, evidence of any harm had a greater influence on increasing acceptance of boundaries than evidence of effectiveness had. In addition, the public perceived outcomes such as death or job loss for individuals in the local community to hold substantial weight and to be more important than inconvenience to a few individuals.
- Evidence of physical harm was the most persuasive factor in accepting boundaries. In most discussions, the preferred way to protect others from harm consisted of guidelines and oversight by medical authorities. In other instances, participants cited and supported rules to prevent adverse effects on the public’s health, such as those now requiring people with tuberculosis to receive treatment.

- Evidence of economic harm was also a persuasive factor in restricting choice. For example, many participants stated that the economic impact and loss of access to care for the community that could result from closing a local hospital, even if it were low performing, outweighed clinical quality for those few who needed specialized surgery. Likewise, participants nearly unanimously supported the need for limits to prevent people from taking advantage of the system and overusing their “fair” share of resources; this was an issue when individual choices increased what others had to pay for health care.

Assessments of risk of physical and economic harm often influenced attitudes about whether society should establish a boundary on decisionmaking: the greater the risk, the more support for the boundary.

- Participants’ perceptions of risk of harm varied, as did the level of comfort with risk-taking. For example, in examining the differences in rates of complications between the low-volume and high-volume hospitals, some participants perceived the level of risk at the low-volume hospital as substantially higher than the risk at the high-volume hospital, while others did not perceive much difference.
- These relative assessments of risk sometimes influenced attitudes about whether society should establish a boundary: the greater the perceived danger, the more support for the boundary.

Although the public believed doctors have the responsibility for knowing and discussing the evidence, they also believed that patients have the responsibility to educate themselves and ask questions of their doctors.

- Participants spoke of doctors’ responsibility to educate themselves about evidence and often identified the doctor as responsible for discussing evidence of benefits and harms with patients so that patients can make informed decisions.
- Most participants believed that patients were responsible for making informed health care choices, asking questions of their doctors, and maintaining a healthy lifestyle. Some strongly supported this perspective from the outset, while others noted that group deliberation changed their views supporting greater patient responsibility.

Doctors—given their understanding of the evidence and the individual patient—should have the authority to determine whether to comply or depart from the evidence in any particular situation. However, the system should hold doctors accountable for their actions to make sure patients receive high-quality health care.

- Participants wanted clinicians to be aware of and generally follow evidence-based guidelines from medical professional associations. Nevertheless, participants believed that clinicians, as experts with specialized education, should be allowed to depart from the guidelines or evidence when needed for individual situations.
- Initially opposed to restricting clinicians’ autonomy, participants often called for increased accountability when faced with evidence that doctors may not always deliver the highest quality care.

Concerns about financial motivations of health care systems, providers, and insurers left many skeptical about whether those setting boundaries or limits in health care would prioritize either evidence of medical effectiveness or quality of care over financial gain.

- Many participants expressed concern that the primary motivation in establishing limits was cost containment rather than ensuring access or quality. Many comments indicated the belief that better care is more expensive and boundaries aimed at cost containment limit access to that better care. Participants were quick to note that costs already constrain patients' choices of and access to certain services.
- Similarly, some participants supported incentives and penalties that could encourage people to adopt healthier lifestyles (e.g., insurance discounts for attending smoking cessation programs) or encourage doctors to provide higher quality care (e.g., professional awards). However, incentivizing physician behavior with financial rewards was more problematic, as participants feared that those incentives might compromise clinicians' integrity by prioritizing financial gain over the patient's health.

The public's trust in entities setting boundaries was influenced by perceptions of expertise, motivation, and whether boundary setting is an appropriate role.

- Overall, participants trusted independent medical associations more than insurers, employers, or government. Participants perceived medical associations as independent, with no financial stake in health care practices or decisions, and as having the needed medical expertise.
- Participants had negative or divided perceptions of other entities based on their perception that such entities lacked medical expertise and/or had questionable motivations. Almost all participants knew that insurers limit care and accepted that as a component of the insurers' role.
- Participants debated whether other payers, such as employers or the government, have the right to set some boundaries. Participants who perceived that these other payers have a legitimate financial or ethical stake in health care tended to accept that these entities could set boundaries. Numerous participants, who had been unaware of the government's large role in paying for health care, became more sympathetic to the idea of government involvement in health care cost containment. Similarly, participants who had been unaware of the risks to society from the overuse of antibiotics tended to become more willing to accept limits on care that promote good antibiotic stewardship.

Throughout deliberation, participants called for more education about evidence and more transparency around health care costs to help inform decisionmaking; some participants even called for government mandates requiring transparent evidence-based information about health care costs, hospital quality, or treatment effectiveness.

- Participants highly supported education and information about health and health care, as most expressed the belief that education and information help people make the best decisions.
- Participants also believed that education about high-quality care is a better approach than restrictive boundaries, especially as education maintains individual freedom and personal choice.

- However, participants held that if education alone is not effective in changing harmful medical practice, then more direct steps for monitoring clinical decisions may be warranted.
- Participants believed that patient access to information about evidence is limited, and a more aggressive effort to bring relevant information to the general public should be a priority. Participants said that the case studies developed for the deliberative discussions would be useful to share with the public: information on provider quality and cost from the hospital quality case study and information on the overuse of antibiotics and MRSA from the URI case study. Participants also wanted general information on treatments and interventions to help improve their decisionmaking.
- Participants noted the difficulties in determining the costs of health care and said that more transparency of health care costs would benefit the public.
- Even though participants generally perceived government interventions that would restrict choice negatively, they typically viewed government mandates requiring transparency, information about costs, and providing evidence-based information about hospital quality and treatment effectiveness positively.

In sum, deliberation required people to consider a variety of tensions and factors in a complex issue, resulting in informed public input that is indepth, nuanced, and actionable. Deliberation allowed participants to explore their own views in more detail, to witness how information and context could influence their perspective and that of others, and to observe how discussion and debate could influence their thinking on the question at hand. As new information or case studies were introduced to the deliberations, answering the overarching question required greater attention to competing priorities. The discussions became more nuanced, with participants exploring the tradeoffs associated with complex individual and societal factors. Although deliberation did not address all misperceptions about evidence or the health care system, numerous participants commented, at the close of their sessions, that they had a deeper understanding of the issues and problems, as well as a better appreciation of a variety of factors relevant to health care.

Effectiveness of Public Deliberation

The randomized design of the Deliberative Methods Demonstration allowed us to assess the impact of deliberation on participants and identify differences by deliberative method and participant characteristics by examining:

- **Changes in participants' knowledge of evidence and comparative effectiveness research.** The knowledge outcome captures the information gained based on questions that were linked to the background educational materials provided to all participants, including the control group. Although participants likely gained additional knowledge from presentations or discussion in deliberative sessions, we measured only the information from the educational materials, which was the most conservative test of increasing knowledge.
- **Shifts in participants' attitudes about the use of evidence in decisionmaking.** Change in attitudes is often measured as an intermediate outcome of effective deliberation. A core assumption of deliberation is that information and discussion may alter the views of

participants as they come to a more informed judgment on the topic. These shifts in attitudes do not have to be for or against a decision; rather, a shift may reflect greater acceptance or greater doubt about one's convictions.¹ Although we used attitude change as a measure of effectiveness, we had no hypotheses for the direction of attitude change. Further, we had no expectation that deliberation would produce group consensus around these attitudes. We assessed attitudes regarding the use of medical evidence in decisionmaking, including questions specific to the hospital quality and URI case studies, and questions on consideration of costs in decisionmaking.

- **Participants' self-reports of the impact the deliberative experience had on them, as well as their assessment of the quality of discourse and implementation.** Impact of deliberative experience included whether participants thought the process affected their views and if participants thought the process was worthwhile. The quality of discourse and implementation included participants' perceptions of the level of participation by all group members, the level of respect for other group members' views, the degree to which participants constructively deliberated the issues, and how well the deliberative methods were implemented.

We assessed these outcomes using two surveys. First, we administered an online survey on knowledge and attitudes to deliberation and control group participants twice, once before educational materials were sent and again within 2 weeks following the conclusion of the deliberative methods. We achieved an 80-percent response rate on the post-survey, using the denominator of all participants recruited ($n = 1,774$). We summarized knowledge scores as a percent of correct answers. After completing a factor analysis using the attitude items, our final attitude measures included six factors and eight single items.

Second, we administered a survey on deliberation quality and experience one time to participants following their participation, either in person or online depending on the deliberative method. Of the 961 participants who took part in deliberation, 878 participants completed the survey, a response rate of 91 percent. After completing a factor analysis of this survey, our final outcome measures included six factors and two single items.

Below, we summarize findings for five research questions addressing the effectiveness of deliberation and summarize per-group implementation costs for each deliberative method. The unit of analysis for Research Questions 1–4 is the individual participant and for Research Question 5 is the deliberative group.

Public Deliberation Compared With Educational Materials Alone

Question 1: Is public deliberation more or less effective than educational materials alone at changing knowledge about the deliberative topic, and is there a concomitant shift in attitudes?

Participating in deliberation increased participants' knowledge of evidence and comparative effectiveness research.

- Deliberation (for members of all groups combined) increased participants' knowledge of medical issues and concepts related to health care in the United States, the use of medical evidence, and comparative effectiveness research as compared to the control group.

In sum, the increase in knowledge in the deliberative versus control groups represents a clear effect of deliberation on information gained and retained above the use of educational materials alone.

Participating in deliberation shifted participants' attitudes regarding the role of evidence in decisionmaking but did not shift views regarding the relative importance of evidence and personal preferences.

- Deliberation (for members of all groups combined) shifted participants' attitudes related to the importance of medical evidence at a statistically significant level, specifically increasing agreement with:
 - The factor *importance of knowing about medical evidence when making health care treatment decisions*
 - The item *medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking*
- A shift did not occur in the factor *doctors and patients should consider evidence over preferences when making treatment decisions*.

In sum, deliberation was associated with a shift from agreement to stronger agreement concerning the role of evidence in decisionmaking. When directly proposed against the role of preferences, participants supported the role for evidence, but deliberation did not change views about the relative importance of evidence versus preferences.

When comparing each deliberative method with the control group, all four deliberative methods showed significant change on at least one knowledge or attitude measure.

- Compared with the control group, the CP and BCD methods increased participants' knowledge about evidence and comparative effectiveness research at a statistically significant level. The CD and ODP methods increased participants' knowledge as well, but not at the level of statistical significance.
- Compared with the control group, each of the four deliberative methods shifted participants' attitudes for at least one measure related to the importance of medical evidence at a statistically significant level. For the CP, CD, and ODP methods, shifts showed increasing agreement with the factor *importance of knowing about medical evidence when making health care treatment decisions*. For the BCD and CD methods, shifts showed increasing agreement with the item *medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking*. For the CP method, shifts also showed increasing agreement with the factor *doctors and patients should consider evidence over preferences when making treatment decisions*.
- Compared with the control group, the CP method shifted participants' attitudes related to considering costs in making treatment decisions at a statistically significant level. Shifts showed increasing agreement with the factor *doctors and patients should consider cost evidence when making decisions*. This factor was evaluated for all methods, as all participants received information on health care costs as context for the discussion. However, the CP method had more time allotted for learning about and discussing issues related to health care costs.

- Attitudes regarding use of medical evidence to restrict antibiotic use reflected a similar impact of deliberation for CP and CD—the two methods that discussed this case study—when each was compared with control. Participants in both methods shifted to more agreement at a statistically significant level on the item *government should limit when doctors can prescribe antibiotics*.

In sum, these findings suggest that all of the deliberative methods can be judged effective compared with a control that used reading materials only on the basis of change on at least one knowledge or attitude measure at the level of statistical significance. However, these statistical tests of individual methods versus control do not allow us to draw conclusions about the relative effectiveness of methods.

Shifts did not occur in three items related to the hospital quality case study, which was used in all methods.

- There were no shifts at a significant level in attitudes related to the material in the hospital quality case study, which all the groups deliberated. This result held true when comparing participants in all deliberative methods compared with the control group, as well as when comparing participants in each method with the control group.

The lack of significant findings may be due to the specific content and complexity of the hospital quality case study. This case study juxtaposed concerns about having access to a “better” high-volume hospital versus the potential impact on the town of having a local low-volume hospital lose business and perhaps close because of reduced patient census. Further, unlike the other case studies, community concerns undermined rather than supported the primacy of evidence.

Overall Quality of Deliberative Sessions

Question 2: What was the overall quality of deliberative discourse and participant experience among the four methods?

Participants reported that they placed a high value on taking part in deliberation and that the experience affected their opinions.

- Participants across all methods placed high value on taking part in deliberation. High ratings of the factor *perceived value of the event* showed that participants valued their participation and included their indication that they would like to participate in activities like this in the future.
- Ratings for the factor *effect of deliberation on participants* reflected participants’ perceptions that the experience had an impact on their opinions on the deliberative topic.

Participants rated the quality of deliberation as high in terms of both the quality of deliberative discourse and the implementation process.

- Participants across all methods rated the quality of communication and discourse highly. Participants reported agreement with the factor measuring the extent that the participants in the groups showed respect for the opinions of others. Participants also reported agreement with the item that people gave reasons to support their opinions. Of note, participants’ ratings for the factor *equal participation in the discussion* were relatively low compared with other measures of discourse quality; participants reported that some people in the group spoke more than others. Despite the fact that participants did not

judge participation to be equal, it did not appear to affect their satisfaction with other aspects of the experience.

- Participants across all methods rated the implementation process highly. Ratings for the factor assessing the quality of the implementation process were overall high, including that the event was well organized, that the information presented was clear and easy to understand, and that the purpose of the event was clear. Ratings for the factor assessing facilitator neutrality were fairly high.

In sum, participants' positive reports of the quality of the deliberative discourse and implementation process indicate that the methods were successful in achieving the core design elements of deliberative methods that were identified in the literature as promoting successful deliberation. Further, positive ratings for the value and effect of deliberation show that participants felt that their input would be used in a meaningful way and that the experience affected them on a personal level.

Comparisons of Specific Deliberative Methods

Question 3: Are specific deliberative methods more effective than others?

Intensity—as measured by contrasting the CP and BCD methods—did not increase knowledge but shifted attitudes at a statistically significant level.

- The higher intensity method (CP) did not increase participants' knowledge of evidence and comparative effectiveness research more than the lower intensity method (BCD).
- Intensity shifted participants' attitudes related to the importance of medical evidence on one factor, *importance of knowing about medical evidence when making health care treatment decisions*, at a statistically significant level. However, intensity did not significantly affect the factor *medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking* or the item *doctors and patients should consider evidence over preferences when making treatment decisions*.
- The higher intensity method (CP) shifted participants' attitudes related to considering health care costs, specifically increasing agreement with the factor *doctors and patients should consider cost evidence when making decisions*, more than the lower intensity method (BCD).

Intensity—as measured by contrasting the CP and BCD methods—had an effect at a statistically significant level on participants' self-reports of the perceived value of the event, the quality of deliberative discourse, and the implementation process.

- Although participants in both methods placed value on taking part in deliberation, participants in the higher intensity method (CP) reported that the experience had a greater impact on them than participants in the lower intensity method (BCD) did. This difference was at a statistically significant level.
- Participants in both methods rated the quality of deliberative discourse and implementation highly, but differed at a statistically significant level for three outcomes:

- Participants in the lower intensity method (BCD) reported more agreement with the two factors measuring equal participation and facilitator neutrality than participants in the higher intensity method (CP) did.
- Participants in the higher intensity method (CP) reported higher ratings of the quality of the implementation process than participants in the lower intensity method (BCD) did.

In sum, intensity of deliberation, as measured by CP and BCD, has marked impacts on shifts in attitudes and resulted in more positive reactions to the impact of deliberation as reported by participants.

Mode—as measured by contrasting the CD and ODP methods— did not change knowledge or attitude at a statistically significant level.

- Our comparison of an in-person (CD) versus online (ODP) method that required a similar total time commitment from participants did not show a statistically significant effect on any of the knowledge or attitude outcomes.

Mode—as measured by contrasting the CD and ODP methods—had an impact on perceptions of the quality of discourse and impact of the deliberative experience.

- Participants in CD reported significantly higher scores than ODP participants for five out of the eight measures of deliberative experience. For the quality of communication and discourse, CD reported higher scores for the factor *respect for the opinions of others* and the item *reasoned justification of ideas*. For the implementation process, CD reported higher scores for the factor *implementation quality*. For participant reports on the impact of the deliberative experience, CD reported higher scores for the factors *effect of deliberation on participants* and *perceived value of the event*.

In sum, remote (online) methods and in-person methods that engage participants for a similar length of time showed similar changes in knowledge and attitude outcomes. However, our comparison showed dramatic differences between the in-person and online methods in deliberative experience, and specifically around perceived value of the event. This result may be due to the particular nature of our online method, in which facilitation was less active. However, remote methods, regardless of facilitation style, may be less likely to inspire the same level of engagement and excitement as in-person methods.

Participants' Personal Characteristics

Question 4: Does the effectiveness of public deliberation vary by participants' personal characteristics?

Deliberation as a method generally affected people from different demographic groups similarly.

- Regardless of race, ethnicity, age, and educational status, participants showed similar increases in knowledge following deliberation.
- The direction and magnitude of the changes in attitude toward using medical evidence in decisionmaking, including mechanisms to support use of high-volume hospitals, were similar across racial, ethnic, age, and educational lines.

In sum, large and consistent differences among groups on knowledge and attitude outcomes would have suggested that deliberation engaged certain demographic groups more or differently than others. In contrast, we observed no differences in changes in knowledge and few differences in changes in attitude outcomes based on demographic group. These findings suggest that deliberation can be equally effective with a wide range of individuals, not just with more educated or privileged members of the public, as has been suggested in the literature.

Participants from historically underrepresented demographic groups may place more value on or perceive greater impact from their participation than others.

- African Americans and Hispanics reported valuing their deliberative experience even higher than others did.
- African Americans and participants with lower educational attainment perceived deliberation as having a greater impact on their opinions than others did.

In sum, these findings further support deliberation as an effective method for getting input from underrepresented populations.

Concordance—the proportion of a group made up of a specific demographic—generally did not affect participant outcomes.

- We found that concordance was not associated with changes in knowledge among our participants from historically underrepresented groups (African-American or Hispanic participants or participants with lower educational attainment).
- Concordance was also not associated with shifts in attitudes about medical evidence, including use of high-volume hospitals, or with the value or effect of deliberation as perceived by participants.
- However, we did find one exception to this result. For African-American participants, concordance (i.e., the proportion of participants in a deliberative group who were also African American) was associated with higher perceived value of deliberation and also with greater attitude change on the factor *people should consider the effect on group premiums when making treatment decisions* (discussed below).

In sum, we found little evidence that group composition (concordance) affects the shifts in knowledge and attitudes that occur in deliberation. Nonetheless, our findings flag the importance of attention to group composition because of selected findings for African-American participants in groups with higher concordance.

Deliberation highlighted or surfaced select content areas in which demographic groups may hold different views.

- All participants moved from disagreement toward neutral on the factor *doctors and patients should consider cost evidence when making treatment decisions* and the item *people should consider the effect on group premiums when making treatment decisions*. However, there were two differences by demographic group:
 - The magnitude of change on both measures was smaller for African-American participants than for other participants at a statistically significant level for both measures. That is, although all participants moderated their views on the appropriateness of considering costs, African Americans were less inclined than

- others to shift this view. (This result controlled for differences in other demographics, including income and education.)
- Elderly participants changed less than others on the single item *people should consider the effect on group premiums when making treatment decisions*.
- Hispanic participants agreed more than others before deliberation that doctors and patients should consider evidence over preferences when making treatment decisions. Following deliberation, Hispanic participants' views moderated and their scores drew closer to those of non-Hispanic participants, but they continued to show more support for consideration of evidence over preferences.

In sum, because there were few differences, we conclude that they do not reflect a differential impact of deliberation as a method. However, they suggest some interesting differences in views, which contribute to our findings on the appropriate use of medical evidence.

Internal Dynamics of Deliberative Groups

Question 5: Do the group-level effects (i.e., the internal group dynamics) of public deliberation vary by deliberative method?

There was little systematic movement toward consensus in the Community Forum groups, and none of the methods systematically reached consensus on any of the three measures we used to evaluate consensus.

- For all three measures, only about half the groups moved toward consensus following deliberation, which suggests that achieving consensus was a random—and not inevitable—process.

We found no evidence that polarization—the systematic tendency of groups and the individuals who compose them to strengthen their predeliberation opinions—occurred among any of the methods.

- Following deliberation, 45 percent of the 1,216 observations, or opportunities for attitudes to move toward the extremes, demonstrated movement away from the midpoint and toward the extremes. Because this rate, or opportunity score, is close to 50 percent, it implies that movement toward the extremes occurred randomly and is not systematic or inevitable. There was also no evidence that some measures were more susceptible to polarization than others.

In sum, small-group distortions that have been reported for jury-like settings were not evident in the deliberative groups. We did not find any systematic patterns of polarization (movement away from the midpoint toward the extremes) or movement toward consensus. These results may offer an argument for designing deliberative methods with the core design features that were held constant across methods in our study: no shared consensus seeking and well-tested and balanced educational materials.

Implementation Costs Associated With Holding Deliberative Sessions

The main costs of deliberation include those of developing materials, recruiting participants, holding sessions, and analyzing and reporting results. The costs we report here are limited to those directly associated with holding deliberative sessions; we exclude additional research-related costs we incurred and some other costs we judged to be difficult to generalize. The implementation costs we report include:

- Participant costs, such as incentives or reimbursement for childcare or transportation
- Facilities costs, such as site rental, food, and drink
- Equipment technology, such as microphones, projectors, Internet connection, and telephone conference lines
- Supplies, such as pens, paper, flipcharts, easels, and markers

Our per-group implementation costs are specific to our approach, including a composition of 12 participants per group in BCD, CD, and ODP, and 24 per group in CP. Per-group costs were:

BCD, \$4,500

ODP, \$4,900

CD, \$6,900

CP, \$23,500

For BCD, CD, and ODP, the largest area of implementation cost that we tracked was that of equipment and technology, accounting for more than half of costs. In contrast, the greatest area of cost for CP was participant-related costs (i.e., incentives, transportation, and childcare).

An important factor affecting the total costs of a deliberative project—not reflected in the costs reported above—is the number of groups typically held when implementing a particular deliberative approach. For example, for a given project, BCD usually convenes 10-12 groups, whereas CP may convene only 1-2 groups.

Discussion and Implications

We highlight implications for the two aims of the project that are relevant to entities that use evidence in decisionmaking, as well as those interested in using deliberative methods.

Our analysis of the public's input into the overarching deliberative question highlighted several areas for those entities that generate, translate, or use evidence to inform decisionmaking:

- Our findings show the public's capacity to apply evidence and view health care issues from a societal perspective—and under certain circumstances, to prioritize societal needs over personal ones.
- Given that participants have particular concerns about the impact of harms—and are willing to accept constraints on their autonomy to address harm—effectiveness studies should be as attentive to this domain as they are to evidence of benefit.

- Researchers and policymakers’ concerns about the known limitations of research evidence are shared by the public. These concerns have implications for generation of evidence and translation of research findings.
- To members of the public, more than to other stakeholders in health care, the term *evidence* covers not only the findings of research studies, but also clinical judgment, test results, trial and error, and common sense. The public’s use and understanding of the term “evidence” highlights the complexity and inherent challenges in efforts to translate and disseminate evidence.
- Supporting the lay public’s use and application of evidence requires more than translating the results of scientific studies into plain language. It also requires that clinical evidence be put in the context of other factors when presented to support personal health decisions, such as values, immediacy of results, convenience, or trust in one’s practitioner.
- The public skepticism about the motivations of insurers, employers, researchers, and government involvement in health care suggests the importance of transparency when it comes to disclosing financial interests in health care overall, and specifically in the generation and use of evidence of medical effectiveness.

Our analysis expanded the evidence base concerning public deliberation methods:

- Deliberative methods offer a feasible and effective approach for organizations to obtain informed public views on complex topics affecting broader constituencies. We found that deliberation had similar effects on people, no matter what their race, ethnicity, age, or educational attainment.
- Our overall assessment was that each method was effective. However, the CP and CD methods may be appropriate for more complex topics, while the BCD and ODP methods may be appropriate for less complex topics. Planners will likely want to consider which types of outcomes are most important, as well as the investment required to implement the deliberative method.
- Planners will likely want to consider which types of outcomes are most important, as well as the investment required to implement the deliberative method.
- Because all methods were effective to some extent in eliciting core values, shifting knowledge and attitudes, and having an impact on participants, our overall findings indicate that there is no one right way to conduct public deliberation. Planners who are developing or modifying methods to suit their needs and preferences can weigh the types of tradeoffs we identify and use our results to inform their choices.

Conclusion

Many organizations—researchers, health care providers, and public and private-sector purchasers—as well as multistakeholder efforts to improve community health have an interest in capturing the public voice on complex and value-laden health issues. Further, multiple topics raised by participants over the course of the Deliberative Methods Demonstration—the financing, structure, delivery, and oversight of health care services—are important policy issues undergoing transformations in concept and design at the local, State, and national levels. The

Community Forum Deliberative Methods Demonstration found that public deliberation was an effective, feasible, and useful method to capture public input on these topics.

For More Information

For more information on the AHRQ Community Forum, please contact Joanna Siegel, Sc.D., in the Center for Outcomes and Evidence at joanna.siegel@ahrq.hhs.gov.

The Community Forum Deliberative Methods Demonstration was conducted by the American Institutes for Research under AHRQ Contract No. 290-2010-00005. Contact Kristin L. Carman, Ph.D., at KCarman@air.org for further information. Organizations participating under subcontract included the Center for Healthcare Decisions, Sacramento, CA, and the Center for Deliberative Democracy and the Center for the Study of Language and Information at Stanford University.

Reference

1. Carman KL, Heeringa JW, Heil SKR, et al. The Use of Public Deliberation in Eliciting Public Input: Findings From a Literature Review. (Prepared by the American Institutes for Research under Contract No. 290-2010-00005.) AHRQ Publication No. 13-EHC070-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2013. www.effectivehealthcare.ahrq.gov.

Section I: Implementation and Research Design

Chapter 1. Introduction

The purpose of the Agency for Healthcare Research and Quality (AHRQ) Community Forum is to improve and expand public and stakeholder engagement in AHRQ's Effective Health Care Program. A primary area of focus for this work is to advance methods for obtaining input from the general public.

As part of this effort, we fielded the Deliberative Methods Demonstration to generate evidence on the effectiveness of using deliberative methods to obtain public input. The demonstration was a randomized controlled trial (RCT) in which deliberative methods were compared with a control group that received reading materials only.

Background on Public Deliberation

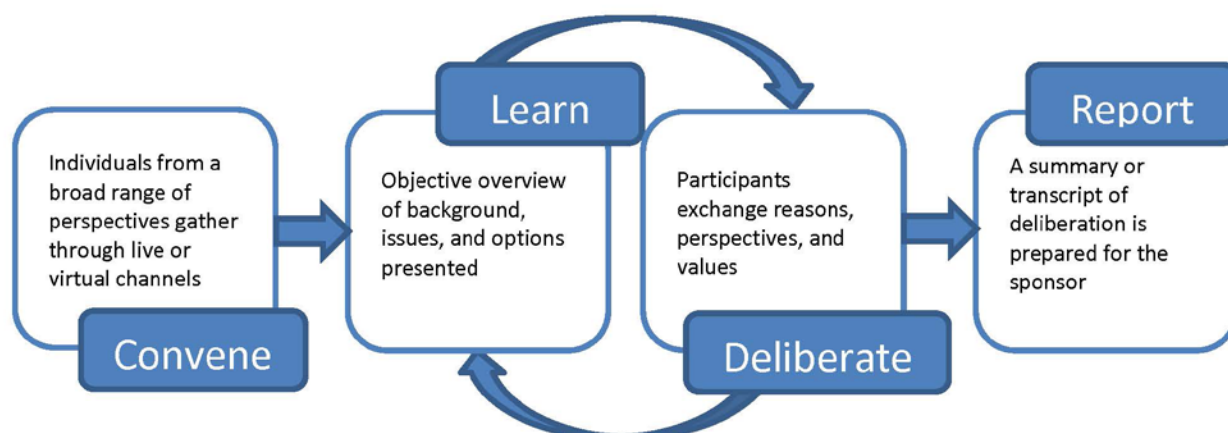
What Is Public Deliberation?

Public deliberation methods provide opportunities to obtain informed perspectives on complex topics that are value laden and that lack simple technical solutions. On such issues, public input on what underlying values should be considered, potential tradeoffs in values, and potential solutions and their likely uptake or resistance are important considerations in developing programs or policies. More information on public deliberation can be found in *The Use of Public Deliberation in Eliciting Public Input: Findings From a Literature Review*, available at www.effectivehealthcare.ahrq.gov/ehc/assets/File/Deliberation-Public-Lit-Review-130213.pdf.

Deliberative methods are a distinct approach to obtaining public input. In public deliberation members of the public are convened to obtain informed input about—and meaningful insights into—how people think about a topic when they are informed. Thus information obtained through public deliberation differs from that collected through surveys or focus groups, which generally obtain more top-of-mind—that is, initial and more intuitive—responses and reactions. In deliberative sessions, participants receive information that is intentionally neutral and respectful of the full range of underlying values, experiences, and possible perspectives. They are encouraged to discuss, learn from others, and examine and refine their own views.

As we show in Exhibit 1-1, public deliberation consists of four core elements: first, convening a group of people (either in person or via online technologies); second, educating the participants on the relevant issue(s) through educational materials and/or the use of content experts; third, having the participants engage in a reason-based discussion, or deliberation, of all sides of the issue(s); and fourth, recording and reporting the dialog to help the sponsors of deliberation incorporate public perspectives when making decisions.

Exhibit 1-1. Core elements of public deliberation



Several distinct deliberative methods have been developed and used previously. They share the four core elements of public deliberation listed above, but differ on key features of implementation such as intensity, whether they take place in person or online, and the use of content experts.

How Has Effectiveness of Public Deliberation Been Evaluated?

Although considerable theoretical and case-study literature endorses the value of public deliberation, little empirical research has been conducted about its effectiveness.¹ In the research that has been done, effectiveness has been defined as—

- **The quality of deliberative experience or discourse.** Based on participant self-reports, researchers' observations, or reviews of session transcripts, these measures typically assess levels of equal participation, active participation, opportunity for adequate discussion, respect for the opinions of others, and awareness of different perspectives.
- **Changes in participants' knowledge or attitudes about the deliberative topic.** A core goal of deliberative methods is informed input, and a core assumption is that information and discussion may alter the views of participants. Thus, typically using pre and post surveys, these measures assess the effect of the deliberation on the participants' knowledge, attitudes, perspectives, values, beliefs, opinions, or policy preferences on the deliberative topics.
- **Changes in participants' empathy and concern for issues affecting the community at large.** Using pre and post surveys, a number of studies have assessed the effect of deliberation on civic engagement and capacity, engagement in the political process, sense of self-efficacy, sense of empowerment, political efficacy and solidarity, and anticipated post-meeting activity related to deliberation issues.
- **Impact on decisions by sponsoring agency.** Ultimately, deliberation obtains information that can influence decisions. Measurement constructs include the effect of public input on specific laws, policies, or practices and on decisionmakers' intentions to act on the results of deliberation. These constructs are usually assessed through case studies or surveys of decisionmakers who may use the findings from the deliberation.

Few well-designed comparative studies of deliberative methods or their alternatives have been conducted. Thus, the evidence base on the effectiveness of one form of deliberation compared with another, or on how various aspects of the deliberative process contribute to outcomes, is limited.

Aims and Study Design

The primary aims of the Community Forum Deliberative Methods Demonstration were to:

- **Inform AHRQ research programs on public views regarding the use of research evidence in health care decisionmaking** by obtaining informed public input on questions central to the mission of AHRQ's research programs regarding appropriate and acceptable ways to use evidence.
- **Expand the evidence base on public deliberation** by evaluating whether public deliberation is an effective and useful way to obtain informed public input for U.S. health care research, as well as to identify a feasible set of choices among deliberative methods.

To achieve these aims, we refined and implemented four distinct approaches to public deliberation on the topic of appropriate use of comparative effectiveness research in decisionmaking. For Aim 1, we conducted a thematic analysis of the transcripts from the deliberative groups to summarize participants' responses to the deliberative question:

Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?

For Aim 2, we constructed a five-arm RCT in which participants were randomly assigned to the four deliberative methods or a reading materials only control group. This experimental design allowed us to assess the impact of deliberation on participants and identify differences by deliberative method or participant characteristics by examining—

- **Changes in participants' knowledge of evidence and comparative effectiveness research.** The knowledge outcome captures the information gained based on questions that were linked to the background educational materials provided to all participants, including the control group. Although participants likely gained additional knowledge from presentations or discussion in deliberative sessions, we measured only the information from the educational materials, which was the most conservative test of increasing knowledge.
- **Shifts in participants' attitudes about the use of evidence in decisionmaking.** Change in attitudes is often assessed as an intermediate outcome of effective deliberation. A core assumption of deliberation is that information and discussion may alter the views of participants as they come to a more informed judgment on the topic. These shifts in attitudes do not have to be for or against a decision; rather, a shift may reflect greater acceptance or greater doubt about one's convictions.¹ Although we used attitude change as a measure of effectiveness, we had no hypotheses for the direction of attitude change. Further, we had no expectation that deliberation would produce group consensus around these attitudes. We assessed attitudes regarding the use of medical evidence in decisionmaking, including questions specific to the case studies on hospital quality and

treatment of upper respiratory infections, and questions on consideration of costs in decisionmaking.

- **Participants' self-reports of the impact the deliberative experience had on them, as well as the quality of discourse and implementation.** Impact of deliberative experience included how the experience affected participants' opinions, whether participants thought the process affected their views, and whether participants thought the process was worthwhile. The quality of discourse and implementation included participants' perceptions of the level of participation by all group members, the level of respect for other group members' views, the degree to which participants constructively deliberated the issues, and how well the deliberative methods were implemented.

The study took place between July 2010 and December 2013.

Organization of This Report

This report is divided into three sections. In Section I, we present details about the deliberative methods design and implementation (Chapter 2) and study design and research methods (Chapter 3). In Section II, we summarize the public's response to the deliberative question (Chapter 4), describe findings from the RCT (Chapters 5–9), and present the results of the cost summary for each deliberative method (Chapter 10). In Section III (Chapter 11), we discuss findings and highlight implications.

Chapter 2. Design and Implementation of the Deliberative Methods Demonstration

This chapter describes the conceptualization and development of the Deliberative Methods Demonstration. We describe the common principles underlying public deliberation that shaped our work, the unique features of the four interventions we used (the deliberative methods), a summary of each of the four methods, the educational materials and case studies used in the deliberative methods, the range of decisions and activities that went into the process of making the Deliberative Methods Demonstration fully operational, and a high-level description of implementing the demonstration. We explain our process and provide detail to support those who may wish to replicate one or more of the deliberative methods in other contexts.

Public Deliberation: Defining Principles and Core Elements

To inform our decisions about the design, measures, and deliberative questions to consider in the study, we conducted a literature review, interviewed key informants, and convened a Technical Expert Panel. (See Appendix A: Technical Expert Panel.) These activities helped us—

- Gain a comprehensive understanding of public deliberation and its underlying principles
- Identify features of deliberative processes that may affect outcomes
- Identify features that affect the feasibility and scalability of public deliberation
- Highlight commonalities and differences among the methods
- Identify methods and measures for our evaluation of public deliberation (Chapter 3)

We assembled a multidisciplinary team of prominent researchers and experts in the fields of public engagement and deliberation, as well as large-scale experimental research design and evaluation, to help design and implement evidence-based deliberation.

What Features Make Public Deliberation Unique?

Public deliberation is a process in which members of the public are faced with an ethical or values-based dilemma and asked to engage in the careful weighing of alternative, often competing, arguments.^{2,3} Participants are often asked to generate solutions or recommendations about an issue for use in a policy or other decisionmaking context.

Public deliberation as a form of public consultation or political participation is an alternative to other approaches, such as opinion polling, that elicit top-of-mind responses. It is also distinct from traditional forms of public engagement, such as focus groups and town halls, because it—

- Emphasizes education and engagement in new information, usually through written materials and sometimes through conversation with experts on the issues at hand
- Requires the practice of reason-giving, as participants are explicitly asked to provide the rationale for their positions
- Demands balance, ensuring that all sides of an issue are considered
- Encourages participants to become social decisionmakers, as well as to consider and speak from their individual points of view.

These features of public deliberation make it uniquely suited for bringing values-based or ethics-laden policy issues to the public for consideration, presenting an opportunity to gain reason-based input from the public on societal issues.

All four of the deliberative methods we chose incorporate these principles. In each group, participants were encouraged, but never required, to view an issue from a social as well as an individual perspective and to consider the tensions around values that were presented, thereby ensuring that participants could express a full range of views.

Also, although some deliberative methods may ask participants to reach consensus, an early and important decision in this study was to avoid requiring consensus to maximize free-ranging discussion.

What Makes for Effective Public Deliberation?

The literature review revealed several design elements that promote successful deliberation. We incorporated them as we designed the demonstration.

Select a topic that affects the common good and to which participants can make a meaningful contribution. During the literature review, we identified and compiled the following criteria to guide selection of the deliberative question for this study:

- An ethics- or values-based issue affecting the common good⁴
- Salient to the public³
- Meaningful (i.e., the input will be valued)⁴
- Providing opportunity for common ground⁵⁻⁷
- Timely and relevant to current public policy^{4,8,9}
- An issue to which the public can contribute^{8,9}

The Technical Expert Panel we convened corroborated these criteria, emphasizing the importance of trust in the process, legitimacy of the process, and assuring participants that their input will make a material contribution to public policy or decisionmaking. This feedback helped move us away from early versions of questions that focused on specific diseases or health services delivery situations. After review and comment by the Technical Expert Panel, the deliberative question addressed larger societal dilemmas related to clinical research and aspects of the comparative effectiveness research enterprise—specifically, the use of evidence. We detail our final selection of the question in the Implementation section later in this chapter.

Include and represent diverse perspectives. Across all methods, participants were selected to reflect the general public, not on the basis of professional affiliation or membership in a stakeholder group. We made special efforts to include African-American women, Hispanic participants, and people age 65 and older. As described in Chapter 3, we developed a complex sampling (and thus recruitment) strategy to match the demographic makeup of a region, inclusion/exclusion criteria, and targets for priority populations.

Provide balanced information as part of the educational component. As a necessary condition for public deliberation and informed public input, participants were given accurate balanced information about the relevant issues. After topic selection, we developed educational materials to provide information on quality health care, medical research and medical evidence,

comparative effectiveness research, and health care costs. We also developed case studies, or vignettes, to illustrate the tradeoffs between competing values and encourage deliberation on the deliberative topic. We used cognitive testing (i.e., in-depth individual interviews) for all written materials to assess comprehension, readability, and perceived balance in presenting information. We created the Community Forum Web site for easy distribution of materials. In addition, we recruited content experts to provide participants with additional information. In the section below, we describe the educational materials and use of content experts in more detail.

Set expectations with participants. We addressed two issues with respect to participant expectations. First, our literature review revealed that deliberation is more successful if sponsoring organizations demonstrate that they value the participants' involvement and will seriously consider their input. To that end, we worked with AHRQ to produce two videos featuring then-AHRQ Director Carolyn Clancy, who conveyed the importance of participation, discussed how input would be used, and provided the visual of a high-ranking representative to make the project more real and exciting. This video was shown at the beginning of each session. A shorter video was shown 1 week prior to the deliberation to help with recruit retention.

The second issue was the need to convey to participants that their life experiences made them experts and that their voices were important, along with the other voices in the group. This framing helped participants feel comfortable speaking up, but also reinforced the need for participants to relate to each other with civility, to respect one another's opinions even when they disagreed, and to give their opinions along with reasons for them.

Encourage active debate and equal opportunities to participate. Regardless of the facilitation methods (active or passive), the ultimate goal was to achieve *active deliberation*, in which the discussion is so engaging that participants stop using the facilitator as an intermediary and simply talk to each other. Tools such as small breakout groups, one-to-one discussions, participant-led discussions, and open space were used in the in-person methods to ensure all participants had the opportunity to engage in the discussion. Other techniques included silence, calling on participants they had not heard from without naming them, and asking for different viewpoints.

Foster a safe environment for reason-giving, open-mindedness, and sharing. Across all groups, participants were asked to engage in reason-giving (substantiating their opinions, thoughts, and preferences with values, ethics, experiences, personal stories, references to data, and other reasons). Our facilitator training, described later in this chapter, addressed techniques for maintaining a safe environment by developing ground rules as a group, encouraging participants to see themselves as experts, and asking participants to share the rationale behind their opinion. We used facilitation tools, such the "five whys," an iterative question-asking technique, to explore underlying values and beliefs.

We carefully considered and operationalized these elements in the development and implementation of all the deliberative methods. Our approaches for addressing each aspect are shown in Exhibit 2-1.

Exhibit 2-1. Approaching the elements that promote deliberation quality and effectiveness

Design Element of Effective Deliberation	Team Approach To Address Element
Select a topic that affects the common good and to which participants can make a meaningful contribution	<ul style="list-style-type: none">• Developed deliberative topic around health research and decisionmaking
Use real-world deliberative methods	<ul style="list-style-type: none">• Selected and designed four deliberative methods based on methods that have been previously implemented
Include and represent diverse perspectives	<ul style="list-style-type: none">• Used a complex sampling strategy to match demographic makeup of region, inclusion/exclusion criteria, and targets for AHRQ priority populations (described in Chapter 3)
Provide balanced information as part of the educational component	<ul style="list-style-type: none">• Developed educational materials and case studies and used cognitive testing to assess comprehension and balance• Recruited experts to provide additional information as needed by method and to capture different sides of relevant issues• Developed Community Forum Web site for easy distribution of materials
Set expectations with participants regarding their role, the activities in which they will participate during deliberation, and how the findings from deliberation will be used	<ul style="list-style-type: none">• Developed and used video from AHRQ, with invitation from the current AHRQ Director• Set expectation that all participants come as experts in their own right
Encourage active debate and equal opportunities to participate in the discussion	<ul style="list-style-type: none">• Conducted facilitator training• Included questions to encourage debate and ground rules for each method in facilitation guides
Foster a safe environment for open-mindedness, sharing, and reason-giving	<ul style="list-style-type: none">• Developed detailed facilitation guides, ice breakers, breakout groups

Note: AHRQ = Agency for Healthcare Research and Quality.

Designing Four Deliberative Methods

We selected four distinct types of deliberative methods that have been used in prior public deliberations and reflect important differences in implementation and intensity. The literature review, the Technical Expert Panel, and our team's experience indicated that such differences should include the number of participants, session length, mode of interaction, and use of content experts. Other characteristics that varied by method included presentation of educational materials, use of breakout groups, communication between sessions among participants, and use of polling during deliberation. The characteristics also had an impact on cost.

As Exhibit 2-2 shows, these methods did indeed vary, featuring different participant time commitments (2–20 hours), principal modes of implementation (online versus in person), and use of content experts, and single versus multiple sessions.

Exhibit 2-2. How the methods varied

Method	Number of Participants Per Group	Structure and Intensity	Mode of Communication	Use of Content Experts
Brief Citizens' Deliberation (BCD)	12 participants	2 hours total; one session	Face to face	None
Community Deliberation (CD)	12 participants	6 hours total; two 2.5-hour sessions 1 week apart with 1 hour average online time between sessions	Face to face; asynchronous online between sessions	Online communication with two experts via discussion board
Online Deliberative Polling[®] (ODP)	12 participants	5 hours total; four 1.25-hour sessions 1 week apart	Teleconference (synchronous online)	Recorded Q&A via a moderator with three experts
Citizens' Panel (CP)	24–30 participants	20 hours total; one 3-day session: two 8-hour days and one 4-hour day	Face to face	Presentations and Q&A with seven experts
Reading Materials Only Group (Control)	N/A	1 hour reading time at home	Written materials received via Web site	N/A

Note: N/A = not applicable; Q&A = question and answer.

Specifically, the deliberative methods and control were as follows:

Brief Citizens' Deliberation (BCD)

In this method, 12 participants met in person once for 2 hours. After participants gave summaries of the educational materials, they discussed a single case study, *Comparing Hospital Quality* (hospital quality), described below. No experts made presentations. A single facilitator and a single note-taker supported these groups. Facilitation was active, designed to encourage attention to the tensions between social values, ethical principles, and the individual versus societal perspective. We held 24 BCD groups, 6 at each location.

Community Deliberation (CD)

This method involved two deliberative sessions, each 2.5 hours long, 1 week apart, for each group of about 12 participants. In the first week, participants discussed the case study *Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment* (URI). During the week between in-person sessions, participants interacted through an online discussion board. In the online setting, two experts provided statements regarding the URI case study, answered participants' questions, and asked questions of their own. At the second in-person session, participants completed discussion of the URI case and went on to discuss the hospital quality case study. A single facilitator and a single note-taker supported these groups. During the in-person sessions, facilitation was intended to be active, as described for BCD above. We held 24 CD groups, 6 at each location.

Online Deliberative Polling[®] (ODP)^a

In this method, each group convened online four times, once a week over a period of 4 weeks, using one case study (hospital quality). Each meeting was a 1.25-hour online session, during which about 12 participants engaged in discussions via a dedicated Web site and Internet-based audio conferencing. Student facilitators with no prior experience in facilitation or health care moderated these groups; they were trained to intervene as little as possible, while still attempting to ensure consideration of the competing arguments in the reading materials. This facilitation style was put in place in order to maintain the neutrality of the moderator. During the first two sessions, participants began exploring issues about hospital quality. Following discussion, the groups had the opportunity to generate questions to be addressed offline by a panel of three experts. The panelists' responses were played back to participants during the third session and were the basis for further conversation in the final session. We held 24 ODP groups.

Citizens' Panel (CP)

Citizens' Panel involved 2.5 days of deliberation: all day on a Friday and a Saturday, plus half a day on Sunday. There were 24 to 30 participants in each group. All five case studies were used. Experts were linked to the group through Skype at key points during the session to provide additional information and different points of view on the issues in the overarching question and the case studies. A clinical expert, who was also a member of the research team, presented on comparative effectiveness research and addressed questions from participants on clinical and related issues. Three facilitators and two note-takers supported these groups. This method permitted the use of smaller breakout groups moderated by a facilitator, as well as an open space in which participants could interact without facilitation. Facilitation in this method was active, designed to encourage attention to the tensions among social values, ethical principles, and the individual versus societal perspective. We held four CP groups, one at each location.

Reading Materials Only Control Group (Control)

Participants assigned to the control intervention received educational materials via an email link. Materials included the same background booklet provided to the deliberative groups, *Preparing for the Community Forum*, as well as three of the case studies: hospital quality, URI, and *Obesity Management: Comparing Treatment and Prevention* (obesity management). We chose three of the five case studies to present to the control as a midway point between the other methods, which received between two and five case studies. Participants did not convene in groups to deliberate. We estimated an hour of reading time.

Commonalities and Differences Among the Methods

The principal commonalities and differences among the deliberative methods are summarized in Exhibit 2-3. As mentioned, we developed our deliberative methods based on their use in the real world, and implemented them as they would be implemented absent the experiment (except for required additional research activities).

^aFindings from this study about the Online Deliberative Polling[®] method do not apply to the In-person Deliberative Polling[®] method.

Exhibit 2-3. Commonalities and differences among the deliberative methods

Commonalities	Differences
<ul style="list-style-type: none"> • Participant makeup: Participants were selected to represent the general public, and not on the basis of their profession, affiliation, “stake,” or membership in a stakeholder group. • Balanced educational materials: Participants were informed on the topic via educational materials that were tested to ensure they are accurate, easy to read and understand, and balanced in their presentation of information and perspectives. • Concrete tasks and examples: Participants gave input on a clearly defined task or dilemma. • Reason-giving: Participants were asked to engage in reason-giving (substantiating their opinions, thoughts, and preferences with values, ethics, personal experiences, information, and other types of reasons). • No goal of consensus: Participants were not asked to reach consensus. 	<ul style="list-style-type: none"> • Length of deliberation: This ranged from 2 hours in the BCD method to 20 total hours in the CP method. • Mode of interaction: Two of the methods used face-to-face deliberation and one method used online deliberation; one method (CD) combined both modes. • Use of experts: Three of the four deliberative methods included experts to help inform participants on the deliberative topic by presenting a range of perspectives demonstrating the complexity of the issue. BCD was the exception. • Review of background materials: Although all groups, including the control group, received complete background materials, BCD and ODP spent less time reviewing these materials during the sessions. • Discussion of case studies: Five case studies were developed. One, the hospital quality case, was used in all methods. A second, the URI case, was used in CD and CP. Three other case studies were used only in CP.

Note: BCD = Brief Citizens’ Deliberation; CD = Community Deliberation; CP = Citizens’ Panel; ODP = Online Deliberative Polling; URI = upper respiratory infection.

Developing the Deliberative Topic

Across all methods, the Deliberative Methods Demonstration elicited public input on the use of research evidence in health care decisionmaking. The topic was presented to participants as a clearly defined dilemma on which to provide input. We posed the following deliberative question to all participants:

Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?

This deliberative topic encompasses several themes, including—

- **Use of evidence to encourage better health care.** Is evidence useful (or what kind of evidence is useful) to a clinician and patient considering a test or treatment that has been found to be ineffective, less effective than another, or riskier than another, or for which effectiveness has not been demonstrated?
- **Use of evidence to encourage better value.** Is evidence useful (or what kind of evidence is useful) to a clinician and patient considering a test or treatment that is effective even though an equally effective but less expensive alternative is available?
- **Decisionmaking when evidence shows more complex tradeoffs.** Is evidence useful (or what kind of evidence is useful) in treatment decisions that involve the balancing of effectiveness, risk, and value?

We determined that the deliberative question was appropriate for deliberation for several reasons. First, the use of evidence in decisionmaking relates directly to AHRQ’s support of research that helps people make more informed decisions and improves the quality of health care services. Public input on this question had the opportunity to make valuable contributions to the AHRQ program. Second, the question required participants not only to understand how evidence

is generated and used, but also to discuss difficult tradeoffs concerning the impact on individuals and communities when evidence is or is not applied in medical decisions. Finally, responses to the question would elicit the public's values about whether patients and physicians have a social responsibility to make evidence-based health care decisions. The question cannot be answered through an examination of technical information alone, because views are inherently shaped by important social values and ethical issues.

Educational Materials and Experts

To inform the development of educational materials, we conducted six focus groups with members of the lay public. These focus groups helped clarify strategies and appropriate content for such topics as comparative effectiveness research, importance of community input, understanding and interest in various medical conditions, and terminology. We also conducted cognitive interviews to assess comprehension and readability of all educational materials.

For the study, we disseminated the educational materials to all recruited participants after they completed the Knowledge and Attitudes Pre-Test Survey (as described in Chapter 3, Research Methods) but before they took part in their deliberative method—about 1 week in advance of their session.

Exhibit 2-4 summarizes the educational materials and experts used for each deliberative method. All methods used the Web site, watched the video featuring Carolyn Clancy, and received the *Preparing for the Community Forum* booklet and the hospital quality case study but varied on other materials.

Exhibit 2-4. Participant educational materials for each deliberative method

Participant Educational Materials	Deliberative Method: BCD	Deliberative Method: CD	Deliberative Method: ODP	Deliberative Method: CP	Control
Community Forum Web site	X	X	X	X	X
<i>Preparing for the Community Forum</i> booklet	X	X	X	X	X
Introductory video featuring Carolyn Clancy	X	X	X	X	X
Case Studies					
Hospital quality	X	X	X	X	X
URI	-	X	-	X	X
Obesity management	-	-	-	X	X
Heart disease	-	-	-	X	-
Fictional case	-	-	-	X	-
Expert-Specific Materials					
Expert photos and biosketches	-	X	X	X	-
Expert PowerPoint presentations	-	-	-	X	-
Expert question and answer recorded session	-	-	X	-	-
Expert statements	-	X	-	-	-

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling; URI = upper respiratory infection.

Background Educational Materials

Prior to their participation, all participants received the *Preparing for the Community Forum* booklet, which described the overall purpose of the project, AHRQ and its reason for sponsoring the project, and what to expect. It also gave definitions and facts on medical research and medical evidence, quality health care, and comparative effectiveness research. Information on rising health care costs and who pays for health care was included to provide context for the discussions. We did not provide information on rules, guidelines, or any other types of boundaries in health care; rather, we allowed interpretations and discussions of boundaries to arise spontaneously. Because of differences in the length of interaction across the four methods, the length of time spent addressing the booklet also varied. The booklet can be found in Appendix B.

Case Studies

To help participants grapple with a complex topic and a fairly abstract question, we developed specific case studies to provide context for each deliberation. The five case studies emphasized certain aspects of the use of evidence and different themes related to the overarching question. Each case study is described briefly here and included in full in Appendix C. We provide more detail on the first case (hospital quality) because it was used in all methods.

Comparing Hospital Quality. This case study presented evidence on the differences in clinical results for several procedures between hospitals that had high versus low volume for such a procedure. All methods used this case study as a way to discuss the overall deliberative question. BCD and ODP discussed only this case study. CD discussed this case study in the second session after discussing the URI case study in the previous session. CP discussed this case study toward the end of the second day of deliberation. The control group also received this case study.

The primary tensions in the case study were—

- Whether evidence of clinical effectiveness should be the sole or primary basis for deciding on a hospital, or whether other aspects of health care, such as proximity to home and family or a better experience of care at a smaller hospital, were just as, or more, important.
- The importance of free choice of provider (hospital) under an insurance plan versus the obligation to follow what the evidence indicated about quality of care to assure better medical outcomes overall.
- Having access to the “better” hospital versus the potential impact on the town of having a local hospital lose business and perhaps close because of reduced patient census.

The BCD and CD methods presented the case study in two separate parts. The first part examined using evidence of clinical results in determining whether individuals should get care at a low- or high-volume hospital. Participants responded to the following prompt statements:

Given the differences in results for some patients in these low- and high-volume hospitals, which of the three statements is closest to your view?

- *How hospital staff treat a person (such as being respectful, communicating well, etc.) is more important than the differences in clinical results.*

- *Regardless of what the evidence shows, people should trust that their doctors will provide the best care, no matter which hospital they go to.*
- *People should only use the hospitals that get the best clinical results.*

The second part of the case study asked participants to assume the role of societal decisionmakers as members of a town council who had to choose between two health plans:

Health Plan A (low-volume hospital). This plan covers all the services available at Springview Community Hospital, the low-volume local hospital. The only services covered for the high-volume Regional Medical Center 50 miles away are specialized care not available at Springview Community Hospital.

Health Plan B (high-volume hospital). This plan covers all services at the larger Regional Medical Center 50 miles away, which has better clinical results for several specialized surgeries. This plan pays for services at Springview Community Hospital for two situations: (1) maternity care and (2) emergency care before transferring patients to Regional Medical Center.

The town council must choose one of these health plans for local government employees. Since the local government is the town's largest employer, choosing Plan B would likely cause the local low-volume hospital to lose many patients and possibly to close. Given this information, participants deliberated about which plan to choose.

The ODP method presented the case study in three parts, including a second part assessing cost considerations with low- and high-volume hospitals. The materials for the ODP groups did not include the prompt statements. Instead, the case study included pro/con tables and a list of discussion questions after each part of the case study.

The CP method included the same information as BCD and CD, but groups discussed the case study as a whole (not broken into parts). Also, this method did not use the facilitator prompts after the first part of the case study, allowing instead for small breakout groups to discuss, with an elected member from the group facilitating, which plan they would choose and why.

Upper Respiratory Infection in Children: Antibiotics Versus Symptom Treatment. Participants in the CD and CP methods, as well as the control group, received this two-part case study. First, participants received evidence that antibiotics, which fight bacterial infection, have no benefits for a child with a viral URI. They were also told that 90 percent of URI cases are viral and that the American Academy of Pediatrics cautions physicians not to use antibiotics unless they know the child has a bacterial infection. Then participants weighed the medical evidence of antibiotic overuse with parent and physician autonomy to use antibiotics (individual choice) and discussed what, if any, boundaries should be established to avoid overuse.

Second, participants learned about methicillin-resistant *Staphylococcus aureus* (MRSA) infection as a societal impact of overusing antibiotics. Participants faced the additional tension of the need to avoid societal harm versus the freedom of choice of individual doctors and/or patients.

Participants in CD discussed the two parts separately during their first session. Participants continued the URI discussion online among themselves and with experts between their two sessions, and readdressed the issues at the beginning of the second session. CP participants

discussed both parts of the URI case study together, focusing on whether the evidence of harm from overusing antibiotics justified societal boundaries.

Obesity Management: Comparing Treatment and Prevention. Participants in the CP and the control group received this study. Multiple social values and ethical principles were in conflict in this case, including, in addition to the primacy of evidence—

- Personal responsibility for lifestyle choices
- The value of prevention versus treatment
- The role of society in shaping an environment that increases people's ability and inclination to eat well and exercise regularly
- The role of government in establishing incentives (i.e., making certain behaviors more costly or difficult)

The case provided background information and evidence about the obesity problem in the United States, health problems associated with obesity, current approaches for managing obesity (e.g., diet, exercise, surgery), and current approaches for preventing obesity (e.g., changing the school or community environments, taxing unhealthy food). The case asked participants to allocate money from a Federal grant to efforts to reduce obesity. The first time we presented this case study to participants in Chicago, we used an open-ended question (without presenting specific options) and found that participants were unsure how to respond. In the remaining three locations, to make the discussion more concrete, participants were asked to choose three of six options to reduce obesity: (1) intensive diet and exercise programs for the moderately obese, (2) bariatric surgery for the moderately obese, (3) changing the community environment, (4) changing the school environment, (5) increasing taxes on non-nutritious foods for people who buy them, and (6) increasing taxes on companies that make non-nutritious food.

Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicine. Used only in the CP method, the heart disease case study provided information and evidence about the relative benefits of using medicines only versus stents plus medicine to treat mild heart disease. This case also involved tensions between two sets of implications: on the one hand, individual patient and physician autonomy and choice, and on the other, the need to protect patients from harm and society from incurring medical expenses that could be viewed as unnecessary.

The evidence in the case indicated that 33 percent of people with mild heart disease should receive a stent as the first line of treatment, but that 55 percent of people with mild heart disease currently receive stents as the first option. The case made clear that patients who had stents also had to take medicines indefinitely. Stents have more risks and are a more costly option but offer faster relief. With this information, participants assumed the role of a patient advisory board to a health insurance plan and discussed what recommendations they should provide to the health plan. The case study was presented on the first day. Participants discussed background information in the large group and deliberated during smaller breakout groups. The full group later reconvened, and a representative from each breakout group summarized its recommendations to the health plan.

Comparing Approaches To Preventing Illness: A Fictional Case. Because the obesity management case study took longer than expected in the first location (Chicago, IL), this case study was used only in the CP method for the final three locations (Durham, NC; Silver Spring, MD; and Sacramento, CA). The case study asked participants to make treatment decisions

regarding a fictional illness called Shake, Rattle, and Roll (SRR). In this case, patients who contract SRR develop severe symptoms and become completely dependent on others for their personal care. The tensions in the case included prioritizing length of life versus quality of life, considering health consequences versus costs, assessing prevention versus treatment, asking people to take personal responsibility (to exercise) versus having society protect everyone (putting medicine in the water), the number of people affected by a problem, fairness, and freedom of choice (having medicine put into the water).

A person would have a 1 in 50 chance of developing this fictional illness. Participants discussed three approaches to managing SRR:

- **Option 1** focused on treating this illness with an expensive medication (\$25,000 per patient per year) that was successful in treating only 50 percent of the cases.
- **Option 2** focused on prevention by encouraging everyone to exercise 30 minutes per day every day. In this option, public health officials estimated that only 50 percent of the population would exercise every day and that it would cost \$1,500 per person to train the public to do the exercises. As such, this was the most expensive option.
- **Option 3** focused on prevention by putting an inexpensive medicine in the water supply (\$5 per person). However, 1 out of every 1,000 people who drank the water would eventually die from an allergic reaction. This was the cheapest option.

Use of Content Experts

Three of the four deliberative methods (CD, ODP, and CP) included experts to help inform participants on the deliberative topic. To select experts, we pursued two strategies recommended by the deliberative literature:

- Retain expertise that provides state-of-the art information that describes contradictions in the information base.
- For value-laden issues, retain expertise on both sides of a position so that participants can become familiar with differing approaches to the content issue.

The experts represented a range of perspectives, demonstrating the complexity of the issue. The three methods used experts in different ways:

- **CD.** Two experts each developed a written statement pertaining to a case study. One was a clinician, the other a clinician with training in bioethics. Their statements were posted on the Web site used by participants between their two in-person meetings. Participants could post questions for the experts and engage in asynchronous discussion with them and each other.
- **ODP.** Three experts responded to questions from ODP participants. Questions were developed in each of the 24 ODP groups separately, then reviewed and consolidated by ODP facilitators into one set of questions, posed to the experts in an audio-recorded panel format. The experts' responses were played back to the participants in their third online deliberative session.
- **CP.** Seven experts presented their views at different times and on different topics via Skype. Participants asked direct questions of the experts, and experts posed questions to

their fellow panelists. One clinical expert attended in person to cover information about comparative effectiveness research, providing an opportunity for questions and answers throughout the session.

Creating the Community Forum Web Site

We worked closely with the Center for the Study of Language and Information at Stanford University to build a custom design based on the Deme content management framework (“Deme sites”). We developed five Deme sites—tailored for each method plus the control—to include the materials for each group. We also used the Web site to give participants written and video materials on the purpose of the project, the sponsoring organization, and what to expect during the group.

Using the Deme Web site to post premeeting educational materials and information about the groups helped us to welcome participants, familiarize them with the process, and provide a personal touch through our facilitator photos and the message from the Director of AHRQ as the first steps in engagement.

Use of Deme for Community Deliberation

We used the Deme platform to support asynchronous communication for CD during the week between in-person sessions. We asked participants and two experts to join an online discussion on the questions and case study from the first meeting, to review an additional case study, and to respond to a poll on the issues discussed during the first meeting. Participants could log in at their convenience during the 1-week period. The experts visited the site regularly to answer participants’ questions, respond to comments, and provide additional perspectives on the topics.

The CD Deme site included the features below. Screenshots of the Deme Web site are shown in Appendix D.

- **Welcome—Home Page and Resources:** Brief description of the Community Deliberation group; links to resources, such as the materials and instructions; tabs for information about the experts; discussion board; and polling Web pages, as well as a resource box with the handouts and materials from the first in-person meeting.
- **Meet the Experts:** Each expert’s biosketch, photo, and brief statement on the deliberative issue(s) being discussed.
- **Discussion Board:** A place for participants to view ongoing discussions, reply to a comment, or start discussion topics (threads). Participants could post their discussion topics and responses at their leisure. All topics and responses were visible only to participants, facilitators, and observers within their CD group.
- **Take the Poll:** A page where participants could agree or disagree with three responses to the issue they discussed during the first in-person meeting. After submitting their responses, participants could post additional views on the statements by commenting in a text box.

Use of Deme for Online Deliberative Polling

We also used Deme to integrate with the WebEx software platform, the primary method for supporting the synchronous communication among ODP participants. Participants signed on to a Web-based teleconference at their assigned time. They used their computers/mobile devices to view the same material onscreen in real time and communicated orally over the telephone conference line. Participants accessed each ODP session via a “Join Discussion” link on their Deme site, which was configured specifically for each of the 24 groups each week. We gave participants contact information for the Deme support team in case they encountered technical issues. Finally, at the end of the 4 weeks, we revealed a tab on each group’s Deme site titled “Take Your Survey,” where participants could complete the Deliberation Quality and Experience survey.

Facilitating the Deliberative Sessions

We developed facilitators’ guides for each method, specifying step-by-step activities, timing, and staff responsibilities for the interactive process. The guides spelled out, sometimes in specific language, what would be discussed, the desired elapsed time for each part of each session, and verbal probes for facilitators to use at will. These probes could refocus the group on various issues in a case or test participants’ response to boundaries by changing factual elements of a case (i.e., “suppose that...”). Individual facilitators could modify language used in the guides so it was consistent with their own style. Changes were not permitted to some language, such as that used in the overarching question.

The facilitators’ guides were developed by the method lead(s) for each deliberative method and reviewed in detail by other team members. Guides were thoroughly reviewed and amended by those who would actually be facilitating. The guides were then tested in nine-person pilots, discussed below, and revised based on participant feedback and observation.

How We Selected Facilitators

Three of the four deliberative methods called for active facilitation, while the fourth method, ODP, used passive facilitation as a way to encourage participants to take a more proactive role. Several of our facilitators were seasoned in public deliberation, while others were experienced in facilitating other kinds of processes (e.g., focus groups) but had little or no experience with deliberation. ODP employed student facilitators with little or no previous facilitation experience; these were selected 1 month prior to the online sessions.

Because we wanted the experiment to be easy to replicate, we chose not to use only experienced public deliberation facilitators in our three in-person methods. Although many people across the country are experienced in focus group or meeting facilitation, far fewer have experience in facilitating public deliberation. The use of facilitators with a range of backgrounds meant facilitator training was part of the project.

How We Trained Facilitators

Training for in-person methods. Facilitator training took place after the facilitators’ guides had been revised based on the pilot findings. For the in-person methods, we held trainings over a 2-day session. All facilitators had a background in health care. Training goals were to introduce the

methods, develop an understanding of deliberation and how it differs from focus groups, create opportunities to practice, engender team building among the facilitation team, and identify outstanding issues and solutions.

Prior to training, facilitators were asked to read the facilitators' guides, case studies, and background information that participants received. They were also given a resource packet, including tips and tools developed by other organizations. As pre-training homework, each facilitator presented a case study to family and friends and led them through a deliberative process. Training focused on practicing skills through mock deliberations, using the facilitators' guide as well as the case studies. Facilitators led small groups through the case studies, developing questions to help the group discuss their values and beliefs on a particular issue. Skilled facilitators were available within each group to provide feedback to new facilitators on presentation skills, strategies for probing, and how to be flexible within groups as they process and communicate.

Trainings were held 3 months prior to the demonstration so that facilitators could continue practicing their skills with friends, family, and coworkers. Bimonthly calls with the full facilitation team let trainees walk through any challenges and enhance facilitators' guides and case study presentations as needed. During the demonstration, facilitators of each in-person method met weekly to discuss their experience, bring up any issues, and provide helpful facilitation techniques to others.

Training for online method. Facilitators for ODP attended a 1-day training. Facilitators were students, the majority of whom had no prior experience in facilitation or in health care, since this method does not require experience in those skills. The goals of the training were to provide an introduction to online deliberative polling, review briefing materials, test online software, create opportunities to practice, and identify outstanding issues and solutions for addressing them. Facilitators were given a facilitators' guide, case studies, discussion points, and a technical guide on how to moderate an online discussion.

During the first half of the day, participants learned how to log into the dedicated Web site and how to start a WebEx session for participants to engage in the discussion. Facilitators had to understand this technology and help guide participants through any technical issues for the first 5 minutes of the session as needed. A technical support specialist was also on the group call to assist after the 5 minutes had passed so the group discussion could begin.

The second half of the day focused on practicing the discussion points laid out for each case study. Students engaged in mock deliberations and led discussion groups by opening the discussion and asking questions from the discussion points document. Students practiced being both a moderator and a participant. The workshop leaders provided real-time feedback to the group, addressing any challenges, and putting an emphasis on objectively moderating the discussion without revealing individual preferences or sharing materials not already in the briefing documents. The training was held 2 days prior to the actual deliberation so that student facilitators retained the information.

Pilot Testing the Methods

We conducted pilot tests of each deliberative method in January 2012, implementing the methods with up to nine community participants. The number of participants ranged from six in CD to nine in CP. Our goal during the pilot was to ensure that the various study components

were implemented effectively, including recruitment and randomization procedures, educational materials, use of experts, technology tools, facilitation, and administration of the Deliberation Quality and Experience Survey, which is described in Chapter 3.

The pilots were facilitated by one lead and one cofacilitator. They used a truncated version of the facilitators' guides, the full background *Preparing for the Community Forum* booklet, and at least one case study. The pilots clearly illustrated changes that needed to be made, including—

- How to introduce the sponsoring organization, which was generally unknown to the participants
- Providing further explanation of why participants' input is valued
- How to frame the hospital quality case study when the tension was not understood
- Changes to the probes to discern participants' underlying values and beliefs.

We also realized in the pilots that the facilitation style was too didactic in the beginning, as facilitators were reading the education materials out loud. We amended our approach by having participants break into small groups at the start of the session to discuss the *Preparing for the Community Forum* booklet, which they then presented to the group.

For the pilots, we used a Chicago-based online recruitment firm. We had a difficult time recruiting and retaining enough participants and ensuring diversity in all the groups. Because of this experience, we decided to use traditional recruitment firms located in the areas in which we were holding the groups. This decision allowed us to better recruit and retain study participants who reflected the local population.

Implementation

The Deliberative Methods Demonstration recruited and randomized 1,774 participants to the four deliberative methods and the control group, convening 76 groups between August and November 2012. After recruitment and assignment to a deliberative method, participants accessed educational materials and logistical information through the Community Forum Web site.

Scheduling and Holding the Groups

The team convened 76 groups in four locations: Chicago, IL; Sacramento, CA; Silver Spring, MD; and Durham, NC. We chose these locations to provide geographic diversity and facilitate recruitment of a diverse sample for this study in terms of racial, ethnic, and sociodemographic background. Sessions were held at a hotel, conference center, or university. Exhibit 2.5 shows the number of groups by month, method, and location.

Exhibit 2-5. Groups by month, method, and location

Method	Chicago, IL, August	Durham, NC, September	Silver Spring, MD, October	Online, ^a October	Sacramento, CA, November	Total Groups Per Method
Brief Citizens' Deliberation	6	6	6	-	6	24
Community Deliberation	6	6	6	-	6	24
Citizens' Panel	1	1	1	-	1	4
Online Deliberative Polling	-	-	-	24	-	24
Total groups	13	13	13	24	13	76

^aOnline groups included participants from all four locations.

Note: All dates were in 2012.

An implementation team consisting of an implementation lead, conference planner, and three onsite managers met weekly to coordinate logistics of all methods—site selection, onsite setup and technical support, materials production, administering participant consent forms, managing the protocol for turning away participants once a group was full, participant and staff logistical support, participant travel and childcare reimbursements, nightly uploading of video and audio recordings from groups, coordination with recruitment and transcription firms, and participant tracking and coordination with the internal recruitment and randomization team.

Debriefing After Sessions

To capture the facilitation teams' immediate impressions and thoughts about the deliberative sessions, we created a standard debrief form that was completed at the end of each session. This form captured the facilitator's and note-taker's impressions of a variety of process factors (e.g., level of equal participation or reason-giving), initial thoughts on the content of discussions, comments about specific shifts in the discussion or the role of specific participants that seemed important, and notes about problems that arose or areas where participants appeared to struggle. Information from these forms was used in several ways:

- To make adjustments during implementation to further refine facilitator questions or develop new approaches to presenting case materials
- To inform changes to the qualitative code list after implementation at the first location (described in Chapter 3)
- To start identifying themes and questions for the summary of public input

Chapter 3. Research Methods

This chapter describes the research questions and methods used to inform Aim 1 and Aim 2. Separate teams studied each aim, but they met weekly to review progress, discuss findings, and integrate findings between aims.

Aim 1: Inform AHRQ Programs on Public Views About the Use of Evidence in Health Care Decisionmaking

To address Aim 1, we conducted a thematic analysis of transcripts from the 76 deliberative groups to summarize participants' response to the overall deliberative question:

Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?

We focused on three main topics related to the overarching deliberative question:

- **Acceptable boundaries.** What, if any, boundaries do participants specify for society's limiting of individuals' and doctors' decisions? Who do participants trust to set boundaries for health care decisions? How do issues of trust (e.g., in research findings, in physicians) influence participants' views on the use of evidence in health care decisionmaking?
- **Situations affecting how participants perceive boundaries.** When considering societal limits on decisionmaking, how do the following situations affect participants' perspectives?
 - Using evidence in decisionmaking when strength of evidence varies (i.e., clear evidence, uncertain evidence, or no evidence)
 - Weighing risks and benefits to the individual and society, including short- versus long-term consequences
 - Assessing individual and societal financial costs for health care options
- **Values and ethics.** What are the values and ethical principles exhibited during the discussion as participants deliberate on societal boundaries and the use of evidence in health care decisionmaking?

This analysis focused on dominant patterns and themes that emerged throughout the discussions. We did not seek to assess changes that occurred over the course of the deliberations, especially changes experienced by individual participants, which we intended to capture with our surveys. When possible, the thematic analysis sought to assess similarities and differences across deliberative methods and case studies. Below, we describe in detail the methods for data collection and analysis.

Data Collection

We audio- and video-recorded each deliberative group. Transcripts were professionally prepared from each audio recording, and video recordings were used to cross-check transcriptions for participant identification and for portions of the discussions where the audio recording was

unclear. In total, we analyzed 157 transcripts from the 76 deliberative groups: 24 transcripts from Brief Citizens' Deliberation (BCD) groups; 48 transcripts from Community Deliberation (CD) sessions; 12 transcripts from Citizens' Panel (CP) sessions; and 73 transcripts from Sessions 1, 2, and 4 of Online Deliberative Polling (ODP) plus a single transcript from the Week 3 plenary session, which was the recording of the experts' responses to questions that all ODP groups viewed during their third meeting. We imported transcripts into NVivo10, a qualitative analysis software program.

Characteristics of the dataset include—

- Dataset size: 2,822,938 words (approximately 6,721 pages, assuming 420 words per page)
- Average transcript length: 17,980 words (43 pages)
- Median transcript length: 17,545 words (42 pages)
- Range of transcript lengths: 5,108–73,222 words (12–174 pages)
- Average percentage of transcript that is facilitator speech: 26 percent
- Number of words that are facilitator speech (actual): 804,942 words (1,916 pages)
- Participants' and experts' speech: 2,017,996 words (4,805 pages)

Data Analysis

We used an iterative process of inductive coding and memo writing and employed a variety of well-established techniques to draw conclusions from the data, such as identifying patterns, assessing the plausibility of findings, and noting relationships between patterns. We tested and confirmed our findings by looking for exceptions and alternative explanations.¹⁰ Next, we describe our process for coding, memo writing, doing an indepth analysis of a small sample of transcripts, and identifying themes for this report.

Coding

Coding is a data-reduction process that assigns tags or labels to segments of text. Codes can then be used to sort data and create output files based on codes or code combinations.

Developing the code list. We developed a list of codes based on the research questions, topics the research team anticipated, and an inductive review of transcripts from pilot deliberative sessions. The inductive review captured topics that participants raised during the pilot session that the research team may not have anticipated.

The code list captured four broad content dimensions related to the overall deliberative question:

- **Actor role within the health care system** covered comments that referred to roles that the various individuals and organizations, such as doctors, insurers, or patients, should or do play in health and health care.
- **Boundary-setting mechanisms** covered comments about ways to set boundaries on health care.
- **Decisionmaking situations** covered comments about factors and processes in health care decisionmaking, such as use of evidence or costs.

- **Values and beliefs** covered comments about different values, beliefs, preferences, and attitudes.

We divided these content dimensions into subcodes for specific topic areas. The codes and subcodes were broad and descriptive, designed for use in matrix analysis. The code list also included deliberative process codes (e.g., reason-giving) and case set coding (e.g., method, location, case study, experts).

Coding the transcripts. We assigned each transcript to one member of the coding team as we received it from the transcription firm between September and December 2012. The coding team consisted of four staff members who were trained to apply codes consistently to the data; three of the coders had also been note-takers during the implementation. Prior to implementation, all coders independently coded a 13-page sample from the pilot transcripts and reached 88 percent agreement using the NVivo tool.^b After coding transcripts from the first geographic location (Chicago, IL), coders revised the code list collaboratively and reached 80 percent agreement using a 14-page sample from another pilot transcript. The team used this final code list for the transcripts from the remaining locations (Appendix E).

Codes were applied at the paragraph level, and coders applied any codes relevant to the content of that paragraph. In addition to the manual coding, transcripts were source-coded for case sets such as method, location, and case study used. Finally, we used query-based coding to identify facilitator speech and to autocode for specific keywords and their surrounding text. The coding team met regularly to discuss coding, how codes should be applied consistently, and how to interpret and code difficult passages. The team also exchanged ideas about patterns emerging from the data and sought alternative explanations for, or exceptions to, emerging themes.

After initial coding was complete, the team further refined and subdivided codes in two ways. We identified two codes that contained a lot of output and could not be easily summarized without further refinement: Boundaries and Values. The team reread and inductively coded output from these larger codes to further refine and subdivide the code. This inductive approach allowed subcodes to emerge from patterns in the output. For example, limits on doctors, boundaries as a slippery slope to stronger restrictions, and boundaries based on medical effectiveness emerged as additional subcodes from the Boundaries code. Next, we identified secondary codes to help organize the data or assess the prevalence of particular attitudes. For instance, the team applied codes that indicated whether participants found boundaries acceptable or not acceptable within the Rules and Regulations subcode.

Memo Writing

Memos are qualitative analysis tools that summarize analysts' ideas about evolving themes and patterns from the data.¹⁰ With a large dataset—as this project had—it is seldom feasible to analyze all coded material. To select a subset of the data for memo writing, we used quantitative techniques to explore the characteristics of the coded data and used the results to help select codes and code combinations for memos.

^bAlthough there are no agreed-upon standards for reliability coefficients, we aimed for 80–90 percent agreement. Krippendorff's standard says that 67–79 percent is acceptable and above 80 percent is good.

Selecting coded data for memo writing. We conducted an exploratory analysis of the coded data to ascertain which codes and code combinations were most prevalent (i.e., occurring frequently) and most pervasive (i.e., occurring in many transcripts); where and to what degree codes overlapped; and whether differences in these code characteristics varied across methods, locations, or case studies. Using multiple NVivo tools, the exploratory analysis included—

- Overall frequency of terms and codes in the data, using term frequency tools (i.e., “tag clouds”) and counts of number of references in the coded output
- Context of high-frequency terms, using the software tool for concordance
- Clustering characteristics of codes and terms in the data or subsets of data (e.g., within methods or cases), using the cluster analysis tools
- Relative magnitude of coded topics within the data, using tree mapping
- Convergence of coded topics, both overall and within methods or cases, using code matrixes

The exploratory analysis identified the following:

- The subcodes within the Boundary-Setting Mechanisms code dimension were the most prevalent, were pervasive across methods and cases, and overlapped with all other content codes in the dataset except Definition of Society.
- Three subcodes within the Decisionmaking dimension—Weighing the Evidence, Weighing Impacts, and Cost—were also prevalent and pervasive, and showed substantial overlap with the Values and Ethics codes.

The team then considered which codes and code combinations would be most useful and highest priority in answering our research questions. The team decided to summarize all material coded with a Boundary-Setting Mechanism code as well as any material coded with the three Decisionmaking subcodes and the Values and Ethics codes with which they overlapped. We also included material from the Definition of Society code, as this code output was fairly small, did not coincide with other codes, and captured discussion related to participants’ understandings of the term “society” that was critical to the research questions.

Using a reductive approach to produce code output. We used a reductive approach to produce code output for the analysts. Specifically, as coded material was extracted from the database for analysis, this material was excluded from subsequent data extraction queries. This approach eliminated the possibility that the same text would be reviewed multiple times if retrieved by multiple queries. This process was repeated as extraction continued, resulting in smaller and more narrowly targeted code output.

All material coded with a Boundary code was reviewed for the boundary memos. Then, material coded for weighing evidence, weighing impacts, and cost was included in the memo writing phase **if**—

- Material was also coded with a Values/Beliefs code
- Material was not also coded with a Boundary code
- Material had not already been retrieved by a previous query

Code output extracted for analysis and memo writing included 618,660 words, or (at 420 words per page) 1,473 pages. This represented 31 percent of all nonfacilitator speech in the dataset and represented 50 percent of coded text.

Writing summary memos of the code output. Six analysts wrote memos summarizing output from 28 codes or code combinations. Each code or code combination was assigned to a single analyst, who used qualitative techniques such as inductive coding, annotations, marginal notes, sketch notes, and highlighting to examine topical themes and patterns in the coded material. Analysts used a standard template featuring —

- A brief description of the code or code combination
- A description of the main themes and patterns found in the code output, including illustrative quotes
- Characteristics of the code or code combination (e.g., number of comments, number of pages of output), including any notable differences by deliberative method or case study
- Observations and questions to help generate meaning from the content.

The analysis team met three to four times per week to share results, which allowed the team to discuss, challenge, debate, and cross-check emerging themes within codes or code combinations and identify emerging overall characteristics and themes across all code summary memos.

The larger project team—including the deliberative method leads, facilitators, note-takers, and quantitative analysts—met in person to review these code summary memos and discuss takeaways and questions related to boundaries and values. After this meeting, the analysts prepared another round of memos to examine relationships across codes and further summarize content. This round of memos included two descriptive memos that summarized key concepts within the deliberative question—evidence, costs, society, and acceptable boundaries; an interpretative memo that identified values and ethics that surfaced through deliberation; and five additional memos summarizing participants’ responses to each of the case studies.

Conducting an In-Depth Analysis, or “Deep Dive”

After finishing the summary memo writing, we conducted an indepth analysis (a “deep dive”) of a selected sample of groups to confirm our findings from the thematic analysis and explore aspects of the quality of discourse we were not able to assess with the thematic analysis.

Selecting groups. We conducted an indepth analysis of a purposive sample of eight deliberative groups, two per method.¹¹ To assess a range of variability within our sample, we selected extreme cases within each deliberative method, those that scored high or low on Factor 1, *importance of knowing about medical evidence when making health care treatment decisions*, from the Knowledge and Attitudes Pre-Test Survey (described under Aim 2 methods). We used this factor because of AHRQ’s interest in understanding attitudes concerning use of evidence and because this factor had the most significant findings of difference between the methods and control group.

Because we did not plan to test differences or make generalizations about methods, we did not keep a variable, such as facilitator or location, constant—or seek representation across these variables—when selecting groups. The selected groups included multiple facilitators. All high-

change groups for the three in-person methods were from California, and no groups from the Illinois location were included in the deep dive.

Summarizing findings from the deep dive. One analyst was assigned to each deliberative method and prepared a memo summarizing main themes and characterizing deliberation for each group that use the deliberative method, as well as similarities and differences across groups. Considering temporal (when the discussion took place, how long it was) and contextual (why the discussion took place, what came right before and after discussions) aspects of discussions, analysts captured high-level summaries of the following:

- Initial discussion of the deliberative question
- Final discussion of the deliberative question
- Attitudes toward evidence
- Comfort with boundaries
- Aspects of quality of discourse, defined in Exhibit 3-1

The analyst met with the deliberative methods leads to obtain feedback on the memo. The team used the memos from each method and notes from the meetings with deliberative leads to summarize similarities and differences across groups and methods.

Exhibit 3-1. What defines the quality of discourse?

Characteristic	Definition
Shifts in perspective	Two types of shifts: (1) A participant shifts perspective from individual to societal or vice versa. (The individual perspective reflects concern for one's own interests or the interests of one's immediate family. The societal perspective reflects a concern for the common good, or the impact of choices and policies on others or the community or society at large.) (2) A participant shifts perspective from expressing concern for self or immediate family to expressing concern for people other than self or immediate family OR expressing empathy for positions other than one's own.
Active debate	Participants discuss with each other an aspect of the case study or the overall question, or a closely related on-topic issue. Debate may be characterized as occurring between participants who largely agree with each other or between participants who hold different views.
Reason-giving	Participants give reasons for their own positions or in support of a position expressed by someone else in the group. Reasons may be directly expressed (e.g., "I think evidence is important because....") or may take the form of illustrative stories (e.g., "Hospital quality is the most important...." followed by a story about a bad hospital experience).

Summarizing Main Themes for This Report

In Chapter 4, we summarize the findings drawn from all memos prepared from analysis of the code output and from the deep dive. The team identified the themes, beliefs, attitudes, and values that were most relevant for answering the research questions and reflecting the public's view on the deliberative question. In addition, we used findings from the indepth analysis to confirm how the process of deliberation appeared to influence participants' responses to the overarching deliberative question. The team of analysts met regularly to debate and challenge the emerging summary, and we returned to the coded data as needed to clarify and confirm our conclusions.

Aim 2: Expand the Evidence Base on Public Deliberation

To address Aim 2, we constructed a five-arm randomized controlled trial (RCT), in which participants were assigned to the four deliberative methods and a reading materials only control group. The four research questions (RQs) for this RCT were—

- RQ1. Is public deliberation more or less effective than educational materials alone at changing knowledge about the deliberative topic, and is there a concomitant shift in attitudes?
- RQ2. What was the overall quality of deliberative discourse and participant experience among the four methods?
- RQ3. Are specific deliberative methods more effective than others?
- RQ4. Does the effectiveness of public deliberation vary by participants' personal characteristics?
- RQ5. Do the group-level effects (i.e., the internal group dynamics) of public deliberation vary by deliberative method?

These research questions address the effect of participation in deliberation on three kinds of outcomes—knowledge of the deliberated topics, attitudes about the deliberated topics, and the experience of participating in public deliberation. Next, we discuss the research questions, participant recruitment and randomization procedures, data collection methods, outcome measures, explanatory variables, and statistical methods. The unit of analysis for RQs 1–4 is the individual participant and for RQ5 is the deliberative group.

Research Questions

RQ1. Is public deliberation more or less effective than educational materials alone at changing knowledge about the deliberative topic, and is there a concomitant shift in attitudes?

RQ1 investigated the impact of deliberation on knowledge and attitudes. It involved comparing deliberation participants with controls to estimate the overall effect of deliberation (i.e., the main effect) and to estimate the effect of the individual methods. Specific questions under RQ1 included—

RQ1.1. Do participants in all four deliberative methods combined demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

RQ1.2. Do participants in each of the four deliberative methods demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

We first pooled deliberation participants from all methods and compared deliberation with the control condition to provide a summary estimate that would answer the dominant question, which is whether or not deliberation is effective. We then compared each deliberative method alone with the control group to establish its individual effectiveness compared with education alone.

RQ2. What was the overall quality of deliberative discourse and participant experience among the four methods?

RQ2 assessed the extent to which we achieved the core requirements of deliberative methods that were identified in the literature as promoting successful deliberation,¹ as evaluated through participants' self-reports of their experience. Specifically, we examined the —

- Quality of deliberative discourse, defined as respect for other perspectives, equal participation in the discussion, giving reasons to support ideas, and having the opportunity to participate
- Quality of the implementation process, including whether the purpose of the event was clearly conveyed, the information presented was clear and balanced, and the facilitator appeared unbiased
- Perceived impact on the participants, including the perceived impact the deliberative event had on the participants in terms of changing their opinions as well as the perceived value of the event

We calculated and reported descriptive means for each measure by method. Higher scores indicated a more positive assessment of the process and impact of deliberation.

RQ3. Are specific deliberative methods more effective than others?

Although the methods shared certain core elements, they varied on a number of features, as described in Chapter 2. RQ3 focused on two features that are particularly important for implementation and that we believed likely to affect individual outcomes—intensity (duration of deliberative sessions and number of sessions required for full participation) and mode of deliberation (the mechanism used for participant interaction during the sessions, either in person or remote). We tested the impact of these two features on knowledge, attitudes, and deliberation quality and experience.

To evaluate the impact of intensity, we compared BCD with CP; to evaluate mode we compared ODP with CD. However, these deliberative methods differ from each other on features other than intensity and mode, so it is not possible to attribute observed differences to intensity or mode unambiguously. We considered systematically varying features of deliberation in constructing our four interventions to provide unambiguous results about the impact of individual features. As noted in Chapter 2, we rejected that approach in favor of studying methods of deliberation that were already in use so that results would be more useful for potential sponsors of public deliberation. Because these comparisons are not solely a comparison of intensity and mode, the results should be interpreted as whole-method comparisons that varied on these key features. As in RQ1, effectiveness was measured by improvement in knowledge, shift in attitudes, and a more positive deliberative experience.

Our specific question regarding intensity was framed as—

RQ3.1. Does the intensity of the deliberative method affect knowledge, attitudes, and deliberation quality and experience?

The deliberative methods varied in length from 2 hours to 20 hours. This analysis compared the outcomes from the method with the greatest intensity (CP) to the group with the least intensity (BCD). The more intense method provided more time for presenting background materials for

greater understanding of the topics and indepth discussion of the issues. It also allowed time for discussion of a broader range of topics. However, the requisite time commitment and associated expense might also make it less desirable or feasible for sponsors, and it could also affect participation by a broad range of the public.

We hypothesized that participants in the most intense method, CP, would experience larger changes in knowledge and attitudes and report a more positive deliberative experience than participants in the least intense method (BCD), because there is more time for thoughtful debate of the topics. Both methods are conducted exclusively in person, so the other priority feature that we thought might affect outcomes, mode of deliberation, was held constant.

The specific question regarding mode of deliberation was framed as—

RQ3.2. Does the mode of deliberation affect knowledge, attitudes, and deliberation quality and experience?

Because of the substantial cost of convening groups in person and the potential to involve people who might not otherwise be able to participate, remote deliberation by telephone and Internet is a potentially popular alternative. However, there is concern that it may be less effective because it is less conducive to the type of participant interaction that is the foundation of deliberation. The impact of mode should ideally be assessed by comparing two methods that are identical except for the use of remote and in-person deliberation. Unfortunately, we were unable to include two such deliberative methods, so we addressed this question by comparing our exclusively remote method (ODP) with the in-person method closest to ODP in intensity (CD). CD involves approximately 6 hours of total participation, while ODP involves 5 hours in total. Although CD is not exclusively in person, it relies mainly on in-person deliberation; it includes two 2.5-hour in-person sessions separated by a week during which participants communicate on a Web site. We preferred comparing ODP with CD to address the question of mode to the alternative of comparing it with our exclusively in-person methods, BCD or CP, because of their substantially different intensity, given our hypothesis that intensity is also an important factor in effectiveness. For RQ3.2, we hypothesized that there would be no measurable difference in change in knowledge, attitudes, or deliberation quality and experience between participants in online (ODP) and in-person (CD) methods of similar duration.

RQ4. Does the effectiveness of public deliberation vary by participants' personal characteristics?

A foundational principle of public deliberation is that all people affected by a policy decision should have equal opportunity to participate in making that decision.¹² To that end, we prioritized diversity and group heterogeneity on race, ethnicity, age, gender, and geography in our recruitment and randomization. The Community Forum Deliberative Methods Demonstration intentionally convened heterogeneous groups of participants to facilitate the exchange of information and ideas among people with diverse backgrounds and perspectives. The racial, ethnic, age, and gender distribution of the participants recruited reflected those distributions in the population of the geographic region in which the deliberative sessions took place. We also slightly oversampled African-American women, Hispanics, and people age 65 and older—members of AHRQ priority populations—to ensure we could estimate effects for these subpopulations.

In RQ4, we evaluated (1) the extent to which we achieved diversity; (2) whether members of underrepresented populations had similar outcomes and experience in deliberation compared with members of majority populations; and (3) whether group concordance, the proportion of participants in the deliberative session who belonged to the same historically underrepresented group as the participant, affected outcomes among underrepresented participants. We explored differences in outcomes by key demographic characteristics (e.g., race, ethnicity, age, education). Specific questions under RQ4 included:

RQ4.1. Do the deliberative outcomes (knowledge, attitudes, perceived impact) of African-American participants, Hispanic participants, participants age 65 and older, and participants without post-secondary education differ from the outcomes of others?

RQ4.2. Does group concordance, or the proportion of participants in the deliberative session who belonged to the same historically underrepresented group, affect deliberation outcomes for African-American participants, Hispanic participants, and participants without post-secondary education?

These research questions determined if participants reflect a broad sample of the local populations from which they were drawn and if group diversity was achieved across all four deliberative methods. In addition, these research questions determined if participants from historically underrepresented populations —

- Have a similar level of change in knowledge and attitudes compared with general population members
- Report similar deliberation quality and experience compared with general population members.
- Are affected by the proportion of their groups who share their personal characteristics

To answer RQ4.1, we pooled data from all participants, regardless of which deliberative method they were assigned to, and compared each underrepresented group of interest with the rest of the sample. For RQ4.2, we constructed a group concordance variable for each participant from a historically underrepresented subpopulation of interest and used it to examine the impact of group concordance on knowledge, attitude, and deliberation quality and experience among those participants.

RQ5. Do the group-level effects (i.e., the internal group dynamics) of public deliberation vary by deliberative method?

Public deliberation is a group activity. Understanding patterns of group behavior and how they differ by deliberative method is, therefore, essential for understanding the relative effectiveness of alternative methods. RQs 1–4 consider effectiveness from the perspective of the individual participant, but RQ5 takes the group perspective. The large number of groups in the Community Forum Deliberative Methods Demonstration—76 among all four deliberative methods—offers a rare opportunity for robust comparisons among deliberative methods on characteristics of group behavior. Two characteristics in particular have been prominently featured in recent literature on public deliberation: polarization and domination. Polarization is the systematic tendency of groups and the individuals who compose them to move “toward a more extreme point in whatever direction is indicated by the members’ pre-deliberation tendency.”¹³ Domination occurs when the deliberative group routinely adopts the opinions of people who are traditionally

more privileged in society—such as men, whites, and people who are better educated—as its predominant conclusions. Polarization and domination are considered weaknesses or limitations of public deliberation in that they work against the intended goal of deliberation, which is to provide policymakers with accurate information about the attitudes held by all segments of the population.

If groups polarize, there is concern that the information provided to sponsors is determined by underlying characteristics of group behavior rather than the content of deliberation, and therefore the information is not actionable. If members of less powerful demographic groups are dominated by members of more powerful groups, there is concern that they are not contributing their unbiased opinions. In the Community Forum Deliberative Methods Demonstration, we used randomization, collection of both pre-deliberation and post-deliberation attitudes, and the large number of groups to examine whether polarization and domination exist in our deliberative methods and compare their prevalence in different methods of deliberation. If polarization and domination are less prevalent in some deliberative methods, it would be an argument for using those methods.

Evidence about polarization and domination primarily comes from research on the behavior of juries, where jurors are compelled to reach a decision based on unanimity or consensus. As noted in Chapter 2, consensus was not discouraged if it occurred in the deliberative methods, but there was no intent to achieve it. The goal of consensus distinguishes jury deliberation from the Community Forum methods and might explain why the inequitable patterns of group behavior observed in juries might not be found in other forms of deliberation. We also examined whether any Community Forum deliberative methods were more likely to reach consensus than others.

To explore these group-level issues, we identified the following RQ5 subquestions:

RQ5.1. Were any deliberative methods more likely to reach consensus than others?

RQ5.2. Does deliberation tend to polarize participants' views?

RQ5.3. Do members of the group from traditionally privileged demographic groups dominate deliberation?

Recruitment and Randomization

In this section, we describe eligibility criteria, recruitment and randomization procedures, the statistical power analysis, and the final sample.

Eligibility criteria. Participants had to meet the following criteria:

- Age 18 or over
- Reside in one of the four geographic locations selected for the study
- Comfortable reading and having conversations in English

- Have Internet access^c to complete online surveys
- Never employed as a practicing physician, nurse practitioner, registered nurse, licensed practical nurse, physician assistant, or health researcher
- Have not participated in one study using interviews or focus groups in the past 6 months or two studies in the past year
- Are willing and available to participate in at least two of the four deliberative methods

Because AHRQ's legislative authorization (Healthcare Research and Quality Act of 1999) directs the Agency to encourage the inclusion of populations that have been historically underrepresented in health care research, we emphasized recruitment of three AHRQ priority populations: people age 65 and older, people of Hispanic ethnicity, and African-American women. As the ODP method requires Internet access, it was necessary to ensure that participants in all five arms had Internet access to avoid having the requirement of Internet access confound comparisons between ODP and other methods.

Recruitment procedure. We recruited 1,774 participants from four locations: Chicago, IL; Sacramento, CA; Durham, NC; and Silver Spring, MD. The study was limited to four locations because three of the deliberative methods require in-person participation at a common location. We selected these locations to achieve racial/ethnic, socioeconomic, and geographic diversity. They include two large, highly urbanized areas with relatively large African-American and Hispanic populations (Chicago and Silver Spring) to ensure adequate samples of ethnic minority populations and two moderate-sized cities surrounded by rural areas within easy driving distance to increase access to suburban and rural residents.

Participants were recruited from existing lists of people interested in research participation by a local commercial recruitment firm at each location to reflect the population's distribution by gender, African-American race, Hispanic ethnicity, and age. Potential participants were screened by telephone or Internet. They received descriptions of the five experimental conditions and indicated which of the four deliberative methods they were willing to participate in, as well as which prescheduled time slots they could attend. The instructions further explained that they would be randomly assigned to one of their selected methods or a control condition in which they would read materials but not take part in deliberation. Recruits had to select at least two of the four methods to be eligible.

Participants who agreed to take part in the study provided informed consent over the telephone at the end of the screening process. Those assigned to in-person methods also signed a consent form at their first group meeting. The American Institutes for Research's Institutional Review Board waived documentation of consent for those who participated exclusively by telephone and Internet, so the verbal consent provided at screening sufficed. Consent covered the deliberative method discussion groups, the Knowledge and Attitudes Pre-Test Survey, the Deliberation Quality and Experience Survey, and audio and video recording. We provide copies of the

^cInternet access was determined in the screening interview by asking recruits: (1) Would you be able to take a survey online? (2) How comfortable do you feel completing a survey online: not very comfortable, somewhat comfortable, or very comfortable? Recruits who answered "no" or "not very comfortable" were deemed ineligible.

telephone screening form and the paper consent form in Appendix F and Appendix G, respectively.

Randomization procedure. The requirement that recruits be willing to participate in at least two study arms, but not necessarily all of them, stemmed from our attempt to avoid a potentially serious threat from selection bias by using “conditional” randomization. In a multitreatment randomized study, it is typical to recruit a sample of participants who are willing and able to participate in any of the treatment options and then randomize the subjects among those options and the control arm. This approach is consistent with maximizing internal validity, a priority for this study. However, a major limitation of the typical randomization approach for this study is the disparate time demand placed on participants by the deliberative methods. Participants who are able and willing to participate in lower burden in-person methods (e.g., BCD) or the online method (OPD) might differ significantly in their personal characteristics (e.g., age) or in some unobserved ways from those who are able and willing to participate in higher burden methods (e.g., CP). We expected the pool of individuals who find all four methods acceptable to be too small and not representative of the local populations, which would introduce concerns about expense of recruiting and poor external validity.

To address these concerns, we limited randomization of each participant to only the methods identified as acceptable during screening. In analysis, we limited pairwise comparisons of methods to only those participants who agreed to participate in both methods when recruited. Because both methods in the pairwise comparison were acceptable to each participant, the threat of selection bias from preference for certain methods was minimized. Exhibit 3-2 displays all 11 combinations of acceptable methods that could have been chosen by participants during screening and the frequency with which each combination was chosen. Participants who selected only one method were excluded, because we would not be able to use those people when comparing one method with another. Although we expected few recruits to be willing to attend all four methods, 34 percent, in fact, were willing.

Exhibit 3-2. Participant selection of deliberative methods during screening: frequency of different combinations

Participants Selecting 2 Methods: 6 Combinations	Participants Selecting 2 Methods: Sample Size N (%)	Participants Selecting 3 Methods: 4 Combinations	Participants Selecting 3 Methods: Sample Size N (%)	Participants Selecting 4 Methods: 1 Combination	Participants Selecting 4 Methods: Sample Size N (%)
ABX	277 (16)	ABCX	391 (22)	ABCDX	593 (34)
ACX	118 (7)	ABDX	170 (10)	-	-
ADX	82 (5)	ACDX	49 (3)	-	-
BCX	21 (1)	BCDX	16 (1)	-	-
BDX	18 (1)	-	-	-	-
CDX	12 (<1)	-	-	-	-

Note: A = Brief Citizens’ Deliberation; B = Community Deliberation; C = Online Deliberative Polling; D = Citizens’ Panel; X = reading materials only control group. Denominator for percentages = 1,747. Note that 27 people were directly recruited to ensure an adequate number of participants or to increase diversity in deliberative groups; their method selection information was not available and is thus excluded from this table.

Using this information from the screener, all recruits were randomly assigned to deliberation or control; then, in a second step, those who were assigned to deliberation were randomly assigned to one of their acceptable methods.

It was necessary to establish a target sample size or quota for each method, because each method has a desirable group size (24 participants in CP; 12 participants in BCD, CD, and ODP). The number of groups in each method was based on the power analysis, which modeled both (1) the number of individuals needed for person-level analysis and the expected intracluster correlation at the group level and (2) the funds and time available to conduct groups. The groups for each method were distributed equally among the four locations. The sampling quota for each method at each location was a function of these factors plus an allowance for attrition from no-shows and dropouts. If the quota for a method at a location had already been met, no additional recruits were assigned to that method; they were randomly assigned to one of the other methods they were interested in. There were always a minimum of two options for random assignment, even when only one acceptable deliberative method remained available, because the control arm was always available.

Finally, random assignment was also predicated on membership in one of AHRQ's priority populations. We required sufficient power to estimate effects for African Americans, Hispanics, and people age 65 and older in each deliberative method. Thus, once all the deliberative methods groups at a location were filled, we assigned the few recruits still needed to achieve the total enrollment quota for the location to the control arm with certainty. Appendix H describes the randomization procedure in detail.

Although complex, this approach helped achieve two of our major analytic goals: random assignment to deliberative methods without creating selection bias and understanding whether deliberation works differently for historically underrepresented populations. The consequence, however, is that participants were assigned to study arms with different probabilities. To compensate, we created a set of weights for use in the analyses that were affected by these unequal probabilities.

The weighting procedure involved calculating a post-stratification weight to be used in the regression analyses that would adjust for imbalance on the following variables: health insurance coverage, income, race, and Hispanic ethnicity. Weighting included the following steps:

- Compare groups on baseline characteristics in order to identify variables causing imbalance between deliberative methods and control group participants.
- Define demographic strata based on variables causing imbalance.
- Calculate population proportion for each stratum (using Census data).
- Calculate sample proportion for each stratum in our sample.
- Calculate weight based on population and sample proportions.
- Fit regression model with the weight.

Appendix H describes the development of the weights in detail. The weight variable was only applied to the RQ1 analyses, which compared deliberative methods overall and individual methods with the control group. The performance of the weight was evaluated by comparing participants' weighted baseline survey scores across the five experimental conditions. The weight resulted in balanced treatment and control groups, as evidenced by no significant differences in baseline knowledge and attitude scores across experimental conditions in the weighted sample. Because RQ2–4 analyses did not use the control group (and the assignment of

participants to a specific deliberative method once they had been assigned to deliberation did not suffer from these demographic imbalances), weights were not used for RQs 2–4.

Sample size. Based on information obtained from previous studies, we estimated sample allocation targets based on the assumption of finding a small to moderate effect for differences in pre-post change scores between experimental and control groups (RQ1). We set the sample size to detect moderate effects for comparison between methods (RQ3), since small differences among methods would be meaningless for practical implementation. Unless differences among methods were at least moderate, sponsors would use the method that was easiest for them to implement.

The literature on effect sizes that can be expected from this design in studies of public deliberation is very limited. Two studies of deliberation measured change in attitudes using multiple outcome measures and found pre-post effect sizes in the 0.17 to 0.69 range.^{14,15} A third study using an RCT compared a post-intervention measure of knowledge between experimental and control arms and found an effect size of 1.67 when comparing in-person deliberation with control and 2.05 when comparing Internet deliberation with control.¹⁶

The power calculations, conducted using Optimal Design Software, were based on these assumptions: (1) the outcome measure is a continuous variable; (2) there are 12 people in each deliberative group; (3) the intraclass correlation is 0.03 within each deliberative group;^d and (4) the Type I error rate is 5 percent for two-tailed tests. Taking these assumptions and the constraints imposed by number of groups and optimal group size into account, we set the total target sample size at 1,296 participants. The target allocation by location and method is displayed in Exhibit 3-3.

^dThe study participants deliberated in groups (i.e., clusters), which is likely to make participants' deliberation experience and attitudes within the same group more similar to each other than they would be if deliberation were an individual endeavor and the participants were independent of each other. We have taken this clustering (i.e., the intraclass correlation coefficient) into account when estimating the minimum detectable effect sizes. Some experts in deliberation argued that we should not take clustering into account, because deliberation is expected to create similarity within each group. However, we designed our interventions to avoid consensus as a goal, although facilitators did not discourage it if it developed naturally. Because consensus was not a goal, accepted statistical practice requires consideration of clustering when individual units of analysis are grouped, and as it is the conservative approach, we adjusted variances for the intraclass correlation coefficient.

Exhibit 3-3. Number of participants targeted, number recruited, and number who showed

Deliberative Method	Number of Participants and Groups: Target Sample	Number of Participants and Groups: Recruited ^a	Number of Participants and Groups: Showed ^a
Brief Citizens' Deliberation			
# participants per group	12	17.2	12.9
# groups	24	24	24
Total sample	288	412	309
Online Deliberative Polling			
# participants per group	12	16	10.9
# groups	24	24	24
Total sample	288	384	262
Community Deliberation			
# participants per group	12	17.9	12.2
# groups	24	24	24
Total sample	288	429	292
Citizens' Panel			
# participants per group	24	33.3	24.5
# groups	4	4	24
Total sample	96	133	98
Deliberative Groups Subtotal			
# groups	76	76	76
Total	960	1,358	961
Control			
Total sample	336	416	377
Total Deliberation Plus Control			
# groups	76	76	76
Total sample	1,296	1,774	1,338

^aAverage number of participants among the groups.

We estimated that, with this sample size and allocation, the minimum detectable effect size (MDES) for the key comparisons would range from 0.21 to 0.40, which is considered small to moderate.¹⁷ We recruited 1,774 people to achieve the desired sample of 1,296 participants. In addition to the target sample allocation for the entire study, Exhibit 3-3 displays the final recruited sample and the number of recruits who attended. Attendance for the control group is defined as taking the pre-intervention survey.

Exhibit 3-4 displays the number of participants recruited and show rates by method, location, and demographic characteristics.

Exhibit 3-4. Participant show rate by method, location, and demographic characteristic

Demographic	Number Recruited, n=1,774	Number of Shows, n=1,338 (Columnar %)	Number of No Shows, n=436 (Columnar %)	Show Rate (%) ^a
Deliberative Method				
BCD	412	309 (23%)	103 (24%)	75
CP	133	98 (7%)	35 (8%)	74
CD	429	292 (22%)	137 (31%)	68
ODP	384	262 (20%)	122 (28%)	68
Control	416	377 (28%)	39 (9%)	91
Location				
Chicago, IL	420	326 (24%)	94 (22%)	78
Durham, NC	420	339 (25%)	81 (19%)	81
Silver Spring, MD	459	327 (24%)	132 (30%)	71
Sacramento, CA	475	346 (26%)	129 (30%)	73
Age				
Under 65 years	1,580	1,175 (88%)	405 (93%)	74
65 years and over	194	163 (12%)	31 (7%)	84
Gender				
Male	797	599 (45%)	198 (45%)	75
Female	977	739 (55%)	238 (55%)	76
Race				
Non-African American	1,274	944 (71%)	330 (76%)	74
African American	498	394 (29%)	104 (24%)	79
Ethnicity				
Non-Hispanic	1,588	1,208 (90%)	380 (87%)	76
Hispanic	186	130 (10%)	56 (13%)	70
Education				
College or post-graduate	842	662 (49%)	180 (41%)	79
Some college, 2-year associate, or vocational	609	454 (34%)	155 (36%)	75
High school/GED or less	325	222 (17%)	103 (24%)	68

^aShow rate calculated by number of shows divided by number recruited.

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Data Collection

In this section, we describe the quantitative data collection methods, including developing the surveys, administering the surveys, and developing the final outcomes measures. The final surveys are shown in Appendixes I and J.

Survey Development

The effectiveness of deliberation for the Community Forum Deliberative Methods Demonstration was evaluated in terms of (1) changes in knowledge, (2) changes in attitudes, and (3) participants' self-rating of the quality of their experience during the deliberative sessions. We measured these constructs using pre-deliberation and post-deliberation surveys of knowledge and attitudes, and a post-deliberation survey of participants' experience. We conducted two rounds of cognitive testing for both surveys with fewer than 10 participants in each round. We also conducted one round of usability testing of the online surveys with fewer than 10 participants to be sure they were programmed correctly.

Knowledge. Since one goal of public deliberation is to increase participants' knowledge about policy topics, we assessed the effectiveness of deliberative methods by measuring change in knowledge, assuming that greater increase in knowledge represented greater effectiveness. We assessed participants' knowledge about issues and concepts described in the background materials (the *Preparing for the Community Forum* booklet) related to health care in the United States, the use of medical evidence, and comparative effectiveness research. Although participants likely gained additional knowledge from presentations or discussion in deliberative sessions, we measured only the information from the educational materials, which was the most conservative test of increasing knowledge.

Attitudes. Since a core assumption of deliberation is that information and discussion can alter the views of participants, we assessed the effectiveness of deliberative methods by measuring change in attitudes about the deliberative topic, the use of research evidence in health care decisionmaking.¹ Although we used attitude change as a measure of effectiveness, we had no hypotheses for the direction of attitude change.

The attitude questions focused on health care decisionmaking when research findings provide no support for, or conflict with, patient and doctor preferences for particular treatments. Specifically, we assessed attitudes toward how results of medical research should be used when making treatment decisions; factors doctors should consider when making treatment decisions; limiting the consumption of high-fat, high-calorie foods; and health care costs.

We also included questions for two of the case studies: the use of evidence to limit antibiotic use and mechanisms to support high-volume hospitals. We included attitude questions only for case studies that were used in more than one method. Further, to reduce respondent burden by minimizing the number of survey questions, the survey questions about high-volume hospitals referred only to evidence about pediatric heart surgery rather than all of the information presented in that complex case study.

Deliberation quality and experience. We developed the Deliberation Quality and Experience Survey to measure participants' perceptions of their participation, elements of the deliberative process, and implementation of the methods. The constructs and domains measured in the survey came from an extensive literature review of evaluation and measurement of public deliberation

methods. From the literature, we identified three overarching domains of experience and more narrowly defined constructs within each domain that are key to determining whether deliberation could be deemed “successful” from the perspective of the participants. The domains were—

- Quality of the deliberative discourse, which tests the stated goals of deliberation and was based on participants’ ratings of how they interacted with other group members during the sessions
- Quality of the implementation process, which indicates how well the deliberative sessions were carried out
- Impact on participants, which measures whether participants thought the process affected their views and thought the process was worthwhile

To develop the survey, we incorporated items from existing measures where available and developed several new items to cover the domains and constructs of interest.

Survey Administration

Knowledge and Attitudes Survey. We administered the survey to deliberative and control participants via the Internet before and after the deliberative sessions. The pre-deliberation surveys were administered immediately following recruitment, while the post-deliberation surveys were administered during the 2- to 3-week period following each method. The post-deliberation surveys for control participants were administered at the same time as the initial round of post-deliberation surveys for participants in the deliberative methods in each geographic location in order to protect against attrition that might result from stretching the administration over a longer period of time.

The pre and post surveys included the same knowledge and attitude questions to allow us to measure change. In addition, the pre survey included the demographic and case mix variables necessary to characterize the study sample, test the success of the randomization, and define population subgroups. We limited our analyses to participants who completed both pre and post surveys. Exhibit 3-5 shows response rates to the pre and post surveys. We calculated the response rate using the denominator of all participants recruited ($n = 1,774$) because all of these participants received the Knowledge and Attitudes Survey, including participants who may have not attended a group or who failed to complete either survey.

Exhibit 3-5. Who responded to the Knowledge and Attitudes Pre and Post Surveys?

Deliberative Method	Number of Participants Recruited	Number Who Completed the Pre-Survey	Number Who Completed Post-Survey
BCD	412	360 (87%)	331 (80%)
CP	133	113 (85%)	112 (84%)
CD	429	363 (85%)	320 (75%)
ODP	384	342 (89%)	296 (77%)
Control	416	377 (91%)	354 (85%)
Total	1774	1,555 (88%)	1,413 (80%)

Note: BCD = Brief Citizens’ Deliberation; CD = Community Deliberation; CP = Citizens’ Panel; ODP = Online Deliberative Polling.

Deliberation Quality and Experience Survey. The survey was administered to participants at the end of the final meeting of their deliberative groups. Participants assigned to in-person methods (BCD, CD, CP) completed paper Deliberation Quality and Experience questionnaires.

ODP participants were directed to a link and asked to complete the survey online immediately following the conclusion of their last ODP session. Exhibit 3-6 shows the response rates for the Deliberation Quality and Experience Survey. We defined as “attended” the people who checked in at registration and attended one or more sessions of a deliberative group. Therefore, the denominator includes 30 late respondents who did not participate in a discussion session, 24 respondents turned away who did not participate in a discussion session, and 907 who actually participated in a discussion session.

Exhibit 3-6. Who responded to the Deliberation Quality and Experience Survey?

Method	Number of Participants Who Attended	Number of Surveys Completed (Response Rate)
BCD	309	280 (91%)
CP	98	96 (98%)
CD	292	262 (90%)
ODP	262	240 (92%)
Total	961	878 (91%)

Note: BCD = Brief Citizens’ Deliberation; CD = Community Deliberation; CP = Citizens’ Panel; ODP = Online Deliberative Polling.

Data Analysis

Measures Construction

Combining sets of multiple items into summary scores can result in more valid and reliable measures of the overall constructs than scores for individual items, and the summary measures are often easier to report and more easily understood.¹⁸ Before we combined items, however, we conducted item functioning analysis by examining the frequency distribution of responses for each item and checking for items with high rates of missing data or items with exceptionally high or low pre-test means (ceiling and floor effects). It was not necessary to drop any items for these reasons. We then recoded items where appropriate so that all the items within the same measurement aspect (i.e., knowledge, attitude, or deliberation quality and experience) were placed on the same metric and a higher score consistently indicated (1) a correct answer for the knowledge items, (2) more importance placed on the role of research evidence or more comfort with societal limits for the attitude items, and (3) higher quality or better experience for the deliberative experience items.

As described below, we used factor analysis for the attitude outcomes and the deliberation quality and experience outcomes—but not for the knowledge outcomes—to define the structure of survey items and determine where it was appropriate to combine items and calculate summary scores for the domains of interest. The factors identified in the factor analyses aligned well with our a priori measurement framework.

We defined constructs within each of the three overarching pre-defined domains of knowledge, attitudes, and deliberation quality and experience.

Knowledge. The eight knowledge questions in the survey were all recoded on the same scale (0 = incorrect, 1= correct) so they could be combined into a composite score in which each item is weighted equally. Five questions had ordinal responses (e.g., “definitely true,” “probably true,” “probably false,” “definitely false,” “don’t know”). For these items, correct answers (whether

“definitely true” or “definitely false”) were coded as 1 and all other responses were coded as incorrect, or 0. Two knowledge questions had multiple correct answers, and respondents could check more than one response. For example, the question, “For a new medicine to be approved for use in the United States, medical research results have to show...” has two correct responses: “The new medicine is effective” and “The new medicine is safe.” For these questions, the responses were coded as 1 if all correct responses were selected. All other combinations of responses were incorrect and coded 0. One knowledge question had multiple choices but only one correct answer. For this question, the responses were coded as 1 only if the respondent selected the correct answer and did not select any of the incorrect responses. One item was ultimately excluded because, on further consideration during analysis, it was decided that there was no indisputably correct answer. The final list of knowledge questions included in our analysis is included in Exhibit 3-7.

We did not use factor analysis to define the underlying structure of the knowledge items because, as expected, the responses to any pair of knowledge items were statistically independent, which means that answering one question correctly did not increase one’s probability of answering another question correctly. Instead, we created a summary knowledge score for each person by calculating the percentage of the seven knowledge items that were answered correctly at each of the two data collection points.

Attitudes. A total of 41 attitude questions addressed the three variations on the primary discussion topic: *Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?* The majority of the attitude questions had responses with a five-point Likert scale (i.e., “disagree strongly,” “disagree,” “neither disagree or agree,” “agree,” “agree strongly” for questions related to the three variations, and “not important at all,” “not important,” “no opinion,” “important,” “very important” for the “primacy of evidence” questions). We reverse-coded 12 items (Item 12 and Items 15–25) so that a higher score indicated stronger affinity for the use of research evidence in decisionmaking, more consideration of cost, or less affinity for individual preferences.

We conducted exploratory factor analysis (EFA) to identify the structure underlying the observed responses. The EFA was conducted on the polychoric correlation matrix using the principle factor method with squared multiple correlations as initial communality estimates and varimax rotation. The number of factors was determined by the eigenvalues and the interpretability of the rotated factor pattern matrix. In addition, we evaluated the internal consistency reliability of each factor by calculating the Cronbach’s alpha value. We used a factor loading cut-point of 0.40 to guide selection of items. However, for borderline items with factor loadings between 0.30 and 0.39, we used reliability analysis to determine if including these items affected the Cronbach’s alpha value. If the Cronbach’s alpha was improved by including an item in the scale, if the item had face validity, and if the item did not load on other factors (suggesting it distinctly defined the pertinent factor), we included the item in the scale. For Cronbach’s alpha, scales with coefficients above 0.70 will provide precise data for use in clinical research in which groups of patients are compared with one another.¹⁸ In this study, scales with Cronbach’s alpha above 0.70 were considered acceptable. One attitude domain, “role of community in limiting unhealthy food,” had an alpha coefficient of only 0.63 but was included because it measured an important substantive attitude.

Because the Kaiser criterion (i.e., dropping all factors with eigenvalues under 1) has a tendency to overextract factors, for the attitude items, we further fit confirmatory factor analysis (CFA) models to explore if there is a more interpretable factor structure with fewer numbers of factors. We conducted this analysis using structural equation modeling (SEM) as implemented by SAS PROC CALIS. The goodness of fit of the model to the data was evaluated using chi-square, the comparative fit index (CFI), the non-normed fit index (NNFI), and the average root mean square residual (RMSEA).

Based on the factor analysis results, a summary score was calculated for each respondent for an identified factor by averaging his or her responses across all the items within a factor. These summary scores were the outcome variables in the statistical analyses used to examine the impact of deliberation. The items included in each factor and Cronbach's alpha scores are included in Exhibit 3-7 (Exhibits 3-7a through 3-7i).

We identified six factors for the attitude items. In addition, we identified eight items that did not load on any of the six factors but were considered important outcomes of interest (shown in bold in Exhibit 3-7). These items were maintained in the analyses as individual items. We decided to eliminate Item 4 from Factor 1 and Items 23 and 25 from Factor 2, because their content was inconsistent with the factors on which they loaded. Since several attitude items related to case studies were not discussed in all the deliberative methods, we further divided the attitude constructs into three categories: outcomes common to all methods, outcomes restricted to CP and CD, and outcomes restricted to CP only. Exhibit 3-7 lists the final set of attitude composites, the items they comprise, and the construct with which they are associated.

Deliberation quality and experience. A total of 33 questions measured different aspects of participants' experience in the deliberations. The questions had three- or four-point Likert scale response options (e.g., "disagree strongly," "disagree," "agree," "agree strongly" or "not at all," "a little bit," "a lot"). Each item was coded from 1 to 4 or 1 to 3 so that higher values indicate higher quality or more positive experience. For scaling, six items were recoded for consistency so that a higher score on the item and resulting scale indicated better experience. The items scored on a 1 to 3 response scale were mathematically transformed to a 1 to 4 scale so that they would be weighted equally in the calculated summary scores.

Although we started with a clearly defined conceptual framework based on the literature for the deliberation quality and experience outcomes, the Deliberation Quality and Experience Survey was a new instrument, so we did not assume we had correctly defined the constructs of interest. Therefore, we relied on the EFA and the reliability analyses guided by theory to determine the final outcomes. Excluding the CFA, we used the same approach to create the deliberation quality and experience scores that we used for the attitude scores. Two deliberation quality and experience domains had alpha coefficients below 0.7 (0.64 for "equal participation in the discussion" and 0.69 for "effect of deliberation on participants"). These domains were included in the final outcome measures because they measured important substantive aspects of deliberation quality and experience. The items included in each factor and Cronbach's alpha scores are shown in Exhibit 3-7.

We initially identified five factors. Two items did not load on the identified factors but were considered sufficiently important to retain as global items. Two additional items did not load on the identified factors and were not sufficiently important to retain, so they were dropped from consideration. Two items were related to the role of experts in the deliberations. Because some

of the methods did not use experts, these items were excluded. This allowed us to calculate common scores across all the methods in order to be able to compare them with each other using a common metric. Exhibit 3-7 lists the final set of deliberation quality and experience composites and items associated with their constructs.

Exhibit 3-7a. Final outcome measures for knowledge of health care issues presented in educational materials^a

Question Number	Composites and Items
KA Q42	The role of medical results in new medicine approval for use in the United States.
KA Q43	The role of comparative effectiveness research in comparing medical treatments.
KA Q44	Payment of health care costs in the United States.
KA Q45	The role of comparative effectiveness research in developing quality standards for treatment.
KA Q46	The effect of unnecessary care on health plan costs.
KA Q47	The effect of expense on medical treatment effectiveness.
KA Q49	Hospitals that perform many hip replacement surgeries each year get better results than hospitals that perform fewer hip surgeries.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

Exhibit 3-7b. Final outcome measures for attitudes toward using medical evidence in health care decisionmaking^{a,b}

Question Number	Composites and Items
Factor 1	Importance of knowing about medical evidence when making health care treatment decisions ($\alpha = 0.87$)
KA Q7	How important is it for you to know the results of medical research when making treatment decisions with your doctor?
KA Q8	How important is it for you to ask your doctor about medical research results related to your health problem?
KA Q9	How important is it for people to know the results of medical research when making treatment decisions with their doctor?
KA Q10	How important is it that people ask their doctors about the results of medical research related to their health problem?
KA Q11	How important is it that doctors tell you about medical research results that are related to your treatments?
KA Q12	Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking.
Factor 2	Doctors and patients should consider evidence over preferences when making treatment decisions ($\alpha = 0.77$)
KA Q15	Doctors should be able to provide a medical treatment even if medical research results show it won't work for the patient.
KA Q16	Doctors should be able to provide a medical treatment even if medical research results show it might not work for the patient.
KA Q17	Doctors should be able to provide a medical treatment they prefer, even if medical research shows that there is another treatment that works better.
KA Q18	People should be able to get a medical treatment they prefer, even if medical research results show it won't work for them.
KA Q19	People should be able to get a medical treatment they prefer, even if medical research results show it might not work for them.
KA Q20	People should be able to get a medical treatment they prefer, even if medical research shows there is another treatment that works better.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7c. Final outcome measures for considerations of costs in health care decisionmaking^{a,b}

Question Number	Composites and Items
Factor 3	Doctors and patients should consider cost evidence when making treatment decisions ($\alpha = 0.87$)
KA Q21	People should be able to get any medical treatment, no matter how much the treatment costs.
KA Q22	Doctors should be able to provide any medical treatment, no matter how much the treatment costs.
KA Q23	For health problems that are not life threatening, doctors should be able to recommend any medical treatment, no matter how much the treatment costs.
KA Q26	People should consider the effect on group premiums when making treatment decisions.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7d. Final outcome measures for attitude toward mechanisms to support use of high-volume hospitals^{a,b}

Question Number	Composites and Items
KA Q37	Insurance companies should require children who need heart surgery to be treated at high-volume hospitals.
KA Q38	Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children.
KA Q39	Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7e. Final outcome measures for attitude toward the use of medical evidence to limit antibiotic use (restricted to Community Deliberation and Citizens' Panel methods)^{a,b}

Question Number	Composites and Items
KA Q34	People should not be able to get antibiotics if it may not help.
KA Q35	Doctors should not be able to prescribe antibiotics if it may not help.
KA Q36	The government should limit when doctors can prescribe antibiotics.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7f. Final outcome measures for attitudes toward the role of society in prevention (restricted to Citizens' Panel method)^{a,b}

Question Number	Composites and Items
Factor 4	Role of community (parents, teachers, and schools) in limiting unhealthy food ($\alpha = 0.63$)
KA Q27	Parents should limit how much unhealthy food their children eat.
KA Q28	Schools should limit how much unhealthy food is served to children.
KA Q29	Teachers should teach their students healthy eating habits.
Factor 5	Role of government policies in limiting unhealthy food ($\alpha = 0.84$)
KA Q30	The government should limit advertising of unhealthy food to children.
KA Q31	The government should tax unhealthy food to discourage people from buying it.
KA Q32	The government should limit the sale of unhealthy food in the US.
KA Q33	The government should ban the sale of food that research has shown to cause heart disease.
Factor 6	Role of government in encouraging (funding/incentives) preventive interventions to encourage healthy behaviors ($\alpha = 0.89$)
KA Q40	People who are obese have more health problems and live shorter lives than people with healthy weights. How much do you support the government's spending money to help obese people lose weight?
KA Q41	How much do you support the government spending money to help people stop smoking?

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7g. Final outcome measures for deliberation quality and experience: quality of communication and discourse^{a,b}

Question Number	Composites and Items
Factor 7	Respect for the opinions of others ($\alpha = 0.81$)
DQE Q1	People in the group respected my opinion.
DQE Q2	People in the group really listened to each other.
DQE Q3	Even when people disagreed, they respected other's opinions.
Factor 8	Equal participation in the discussion ($\alpha = 0.64$)
DQE Q5	Some people in the group spoke a lot more than others.
DQE Q6	Some people in the group spoke too much.
DQE Q7	Some people in the group barely spoke at all.
DQE Q8	Speaking as much as wanted to in group.
DQE Q11	Reasoned justification of ideas.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7h. Final outcome measures for deliberation quality and experience: implementation process^{a,b}

Question Number	Composites and Items
Factor 9	Implementation quality ($\alpha = 0.83$)
DQE Q21	The event was well organized.
DQE Q22	There was enough time to fully discuss all the relevant issues.
DQE Q23	I understood what I was asked to do throughout the discussion group.
DQE Q26	The facilitator made sure that all opinions were considered.
DQE Q27	The facilitator encouraged everyone to have his or her say.
DQE Q28	The information presented was clear and easy to understand.
DQE Q29	There was so much information it was difficult to absorb it all (reverse coded).
DQE Q30	I understood the issues well enough to participate as much as I wanted to.
DQE Q31	The materials presented all sides of the issues.
DQE Q33	The purpose of the event was clear.
Factor 10	Facilitator neutrality ($\alpha = 0.82$).
DQE Q24	The facilitator sometimes tried to influence the group with his or her own opinions (reverse coded).
DQE Q25	I could tell that the facilitator favored some positions over others (reverse coded).

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Exhibit 3-7i. Final outcome measures for deliberation quality and experience: participant ratings of the impact of deliberation^{a,b}

Question Number	Composites and Items
Factor 11	Effect of deliberation on participants ($\alpha = 0.69$)
DQE Q10	The discussion led me to change some of my opinions.
DQE Q15	How much did the group discussions affect your opinions on the issues?
DQE Q16	How much did the educational materials affect your opinions on the issues?
DQE Q18	How did you feel during the group discussions?
DQE Q19	How did you feel during the presentations?
DQE Q20	I think the recommendations made by the group will influence how research findings are used.
Factor 12	Perceived value of the event ($\alpha = 0.78$)
DQE Q12	Attending the discussion today was good use of my time.
DQE Q13	I would participate in activities like this in the future.
DQE Q14	More events like this should be held as a way of getting the views of people throughout the country.

^aPlease see Appendix I for the full Knowledge and Attitudes questionnaire and Appendix J for the full Deliberation Quality and Experience questionnaire.

^bBolded rows indicate factors or individual items used as primary outcomes in the analysis. Nonbolded rows indicate the individual items that comprise the bolded factors.

Analytic Models

We used the multiple regression approach for knowledge and attitude outcomes and the general estimating equations (GEE) approach for deliberation quality and experience outcomes (adjusting for correlation of participant responses within groups) in SAS to answer the major research questions. For all research questions, we measured relative rather than absolute changes in attitudes, because relative change allowed us to identify the direction participants' attitudes shifted.

To answer RQ1 (*Is public deliberation more or less effective than educational materials alone at changing knowledge about the deliberative topic, and is there a concomitant shift in attitudes?*) we evaluated whether mean pre-post changes of the knowledge and attitude outcomes were the same between all deliberative methods combined and the control group. To answer RQ2 (*What was the overall quality of deliberative discourse and participant experience among the four methods?*) we calculated the means for each deliberation quality and experience measure by method. To answer RQ3 (*Are specific deliberative methods more effective than others?*) we investigated whether intensity and mode of deliberation affect deliberation effectiveness by comparing the method with the greatest intensity to the method with the least intensity and the exclusively remote method to an in-person method of similar intensity. We compared them on mean pre-post changes of the knowledge and attitude outcomes and the means of deliberation quality and experience outcomes. To answer RQ4 (*Does the effectiveness of public deliberation vary by participants' personal characteristics?*) we evaluated (1) the extent to which we achieved diversity; (2) whether participants from historically underrepresented populations had a similar level of changes of knowledge and attitude outcomes as well as similar experience in deliberation compared with the rest of participants; and (3) whether the proportion of participants from underrepresented populations in a group (i.e., group concordance) affected outcomes. To answer RQ5 (*Do the group-level effects [i.e., the internal group dynamics] of public deliberation vary by deliberative method?*) we computed values for several group process constructs that are commonly used in the public deliberation literature. These constructs include consensus,

polarization, and domination. The analysis determined if selected attitude outcomes exhibited consensus, polarization, or domination at the group level.

Because the participant sample and analysis models varied among the five RQs, more details about the specific analytic approach and models are in Chapters 5–9, where we report results for the RQs.

Strengths and Challenges of the Community Forum Deliberative Methods Demonstration

The Community Forum Deliberative Methods Demonstration contributes to the research history and literature in two areas of growing interest: (1) public knowledge and attitudes about the complex tradeoffs between the use of evidence from comparative effectiveness research and individual latitude when making health care decisions and (2) public deliberation as a method to obtain informed public opinion about important public policy issues in general and the use of medical evidence in particular. Our team of investigators confronted a number of significant challenges concerning the content of the deliberations, the methods of deliberation used, and the research design. In this section, we summarize the choices we made to address these challenges to document the often competing factors we considered, clarify our reasoning, and provide future public deliberation researchers with approaches to consider in their work.

Content of Deliberation

AHRQ established the role of comparative effectiveness research in making individual and public decisions about health and health care as the topic of deliberation from the outset of the project. To refine this topic, we reviewed the literature and consulted with experts on medical decisionmaking and public deliberation to develop educational materials and five detailed case studies to help participants grapple with an abstract deliberation question. The case studies were carefully designed to emphasize different themes and tensions within the overarching deliberative question about the use of evidence in decisionmaking. To provide a fair test of deliberation and minimize the risk that our findings could be confounded by issues related to the design of the intervention, we strongly emphasized the development of the educational materials and case studies we used in our deliberative groups. Both were subjected to extensive cognitive testing for readability and comprehension, far greater than the level of testing generally used to prepare such materials based on the type of testing documented for other deliberative projects in the literature.

Deliberative Methods

We selected and refined four distinct types of deliberative methods used in the Community Forum Deliberative Methods Demonstration because they had been used before with anecdotal success. Each was implemented as a whole, with fidelity to the main procedures, so that future sponsors of public deliberation could easily choose and implement the method most appropriate for their needs based on our findings. Each method was tested and refined in a pilot test to minimize the risk that our findings could be confounded by issues related to the design of the intervention.

We then faced the challenge of organizing, recruiting participants for, and conducting 76 deliberative groups in 4 months. It is unusual for a group-level social experiment to have this many groups, especially when a facilitator and assistant must be present at every session and technical experts at many of them. We completed the 76 groups on schedule. The decision to establish a dedicated implementation team turned out to be essential to the success of the project. Equally important were the weekly discussions between the implementation and evaluation teams to ensure we had the strongest possible research design, given the demands of the deliberative methods. For instance, the 2.5-hour BCD groups could not deliberate the same number of case studies as the 2.5-day CP groups. Working together, the teams arrived at a solution that permitted BCD groups to deliberate one case study topic and the other methods to deliberate more case studies. However, the one case study deliberated by the BCD groups was also deliberated by every other group to anchor the deliberations around a common topic.

Research Approach

Use of a Randomized Controlled Trial

Using RCTs in social science research is inherently difficult, as willingness to participate often depends on which arm the participant is assigned to. Nevertheless, because RCTs mitigate a number of threats to internal validity more effectively than other research designs, we decided at the beginning of the project to use this approach and adapt it as necessary. Two challenges to randomization required adaptation.

The first challenge involved unequal probability of assignment to study arms, since we gave priority to assigning the priority population members to deliberation early in recruitment. We did this to ensure we had enough sample members to determine if race, ethnicity, and age are associated with the effectiveness of each method. This meant that priority population members were more likely to be in the deliberation arms than the control arm. We addressed this issue by using sampling weights for all comparisons of deliberative methods with the control group, which is a standard approach to compensate for unequal distributions.

The second challenge, selection bias from differences in willingness to participate, was addressed with an innovative approach to randomization. Typical multitreatment randomized studies minimize selection bias by recruiting participants who are willing and able to participate in any of the treatment options and then randomizing the subjects among those options and the control arm. This procedure is assumed to equate the study arms on unobserved characteristics. As described in our recruitment and randomization approach earlier in this chapter, such a randomization approach for the Community Forum would have required screening for a small atypical group of participants who were willing to participate in all five arms, which would have been expensive and limited external validity. To address this concern, we limited eligibility to recruits willing to participate in at least two study arms, and randomization of each participant was limited to only the methods identified as acceptable during screening plus the control arm. In analysis, we limited pairwise comparisons of methods to only those participants who had indicated that both methods being compared were acceptable. Systematic variation among intervention arms in willingness to participate is a significant challenge for randomization in social experiments, because the interventions often impose different levels of burden on participants. This innovation equates the samples being compared on willingness to participate as a way of minimizing bias from unobserved factors associated with willingness.

Rapid-Cycle Improvement

Traditional evaluation designs assume that the interventions are constant throughout the implementation period. However, it is now widely acknowledged that, in a rapidly changing environment, it makes little sense to preserve the intervention in its initial form if work suggests clear ways to improve it. There is a tradeoff between increased heterogeneity of the intervention and improved intervention design and probability of success. In the Community Forum Deliberative Methods Demonstration, we conducted group sessions consecutively at four locations over a 4-month period, which gave us the opportunity to improve several elements of the interventions following the first location. We did not alter the fundamental characteristics of the deliberative methods, but we made subtle changes to the way we handled groups when too few or too many participants showed or when participants arrived late, and we altered some elements of the educational materials that confused participants. We proposed integrating rapid-cycle improvement into formal evaluation in this way as a potential model for future investigations of social interventions because of the latitude it offers for improving the intervention over time without sacrificing the analytic design.

Qualitative Analysis Approach

The team applied rigorous standards to the qualitative analysis, using an iterative process of reducing the data, displaying the data, identifying themes and patterns, and verifying conclusions. We protected against bias by testing and confirming our findings through examining exceptions and testing explanations. Given the large volume of qualitative data collected during the study, the qualitative analysis focused on the priority question of dominant patterns and themes of all data related to boundary-setting mechanisms. We did not comprehensively analyze full transcripts or explore temporal shifts in knowledge and attitudes or demographic subgroups, which limited our ability to assess trends by deliberative method and participant inconsistencies. For example, if we conducted a full analysis of all material coded with individual patient role and responsibility, we could determine if personal responsibility was valued more highly than our current analysis suggests. Also, a more indepth analysis might better explain differences between deliberative methods on the Deliberation Quality and Experience Survey and Knowledge and Attitudes Survey.

Sustainability

Long-term sustainability is rarely addressed in the evaluation of social interventions, since funding for the evaluation usually ends shortly after the intervention. Short-term sustainability, however, can be measured if the research design accounts for it. In the Community Forum Deliberative Methods Demonstration, we waited 1–3 weeks after each participant's final deliberation session to obtain responses to the post-deliberation survey. We intentionally delayed collecting the post-deliberation survey data to see if knowledge and attitude changes persisted.

External Validity

As in most RCTs, participants were recruited based on an a priori set of eligibility criteria designed to meet the needs of the study. This approach gives priority to internal validity at the expense of external validity or generalizability. Because implementation staff had to be present physically for all sessions (except for ODP, which was remote) we were unable to sample participants from across the country. We compensated for this situation with two solutions.

First, we intentionally included locations from different regions that were heavily urban (Silver Spring, MD, and Chicago, IL) or had small cities easily accessible from surrounding rural areas (Raleigh-Durham-Chapel Hill, NC, and Sacramento, CA). This enabled us to recruit participants from the types of locations in which the vast majority of Americans live and obtain a sample that varied widely on geographic and demographic characteristics. Within each of the four locations, we recruited a sample that matched the U.S. Census distribution of the entire adult population in these four locations on several demographic characteristics.

We excluded people who had no access to the Internet, because the ODP method required Internet participation and the CD method offered Internet communication between in-person meetings as an option for participants. To assure the validity of comparisons with methods that do not use the Internet, we excluded people without Internet access from those other methods also. We accepted people with limited access to the Internet, so the sample included people with very limited computer and Internet skills.

Summary

The intervention design and the research design were both developed to address significant scientific and logistical challenges associated with this study and involved input from a broad team of researchers. This summary has described the rationale for a number of design choices that ultimately served to improve the quality and enhance the rigor of our interventions, study design, and analysis. Some of these design approaches may be applied to future studies of deliberative methods or other social science research.

Section II: Research Findings

Chapter 4. Answering the Deliberative Question: The Public's Perspective

To inform AHRQ research programs on public views regarding the use of research evidence in health care decisionmaking, this chapter summarizes the thematic analysis of how participants from the 76 deliberative groups responded to the overarching deliberative question:

Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?

The research questions for the thematic analysis focused on three main topics related to the overarching deliberative question:

- **Acceptable boundaries.** What, if any, circumstances do participants specify for society's restricting the decisionmaking of individuals and doctors? Whom do participants trust to set restrictions on health care decisions? How do issues of trust (e.g., in research findings, in physicians) influence participants' views on the use of evidence in health care decisionmaking?
- **Situations affecting how participants perceive those circumstances.** When considering societal limits on decisionmaking, how do the following situations affect participants' perspectives?
 - Using evidence in decisionmaking when strength of evidence varies (i.e., clear evidence, uncertain evidence, or no evidence)
 - Weighing risks and benefits to the individual and society, including short-term versus long-term consequences
 - Assessing individual and societal financial costs for health care options
- **Values and ethics.** What are the social values and ethical principles exhibited during deliberation on societal restrictions and the use of evidence in health care decisionmaking?

In this chapter, we first describe the sample of participants who attended the deliberative sessions and describe the case studies to provide context for participant responses. Then we present findings on the participants' responses to the deliberative question and highlight values elicited during the deliberations. We conclude with a brief discussion.

Background

As discussed in Chapters 2 and 3, we summarized qualitative data from 76 deliberative groups from four distinct deliberative methods. Brief Citizens' Deliberation (BCD) had 24 groups with 281 participants. Citizens' Panel (CP) had four groups with 97 participants. Community Deliberation (CD) had 24 groups with 269 participants. Online Deliberative Polling (ODP) had 24 groups with 260 participants. Exhibit 4-1 details the demographic information of participants by method and across methods.

Exhibit 4-1. Demographic characteristics of participants in deliberation

Characteristic	BCD (n = 281)	CP (n = 97)	CD (n = 269)	ODP (n = 260)	Total (n = 907)
Location					
Chicago, IL	72 (26%)	23 (24%)	66 (25%)	55 (21%)	216 (24%)
Durham, NC	68 (24%)	25 (26%)	68 (25%)	75 (29%)	236 (26%)
Silver Spring, MD	69 (25%)	28 (29%)	66 (25%)	63 (24%)	226 (25%)
Sacramento, CA	72 (26%)	21 (22%)	69 (26%)	67 (26%)	229 (25%)
Gender					
Female	157 (56%)	55 (57%)	148 (55%)	135 (52%)	495 (55%)
Male	124 (44%)	42 (43%)	121 (45%)	125 (48%)	412 (45%)
Age					
Under 65 years	248 (85%)	85 (88%)	235 (87%)	239 (92%)	797 (88%)
65 years and over	43 (15%)	12 (12%)	34 (13%)	21 (08%)	110 (12%)
Race/Ethnicity					
White	166 (59%)	41 (42%)	149 (55%)	159 (61%)	515 (57%)
Black or African American	72 (26%)	45 (46%)	85 (32%)	64 (25%)	266 (29%)
Other race	43 (15%)	11 (11%)	35 (13%)	37 (14%)	126 (14%)
Hispanic (any race)	38 (14%)	10 (10%)	31 (12%)	32 (12%)	111 (12%)
Education^a					
High school or less than high school	73 (26%)	25 (26%)	59 (22%)	49 (19%)	206 (23%)
Some college	79 (28%)	33 (34%)	83 (31%)	76 (29%)	271 (30%)
College degree	129 (46%)	39 (40%)	127 (47%)	135 (52%)	430 (47%)

^aTotal population 25 and older.

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

To help participants grapple with a complex deliberative question, all participants received the *Preparing for the Community Forum* booklet, which gave definitions and facts on medical research and medical evidence, quality health care, and comparative effectiveness research. Information on rising health care costs and who pays for health care was included to provide context for the discussions. We did not provide information on boundaries in health care; rather, we allowed interpretations and discussions of boundaries to arise spontaneously. We also used case studies based on specific examples of comparative effectiveness research to illustrate the issues. The case studies are summarized in Exhibit 4-2 to provide context here for participant responses. Appendix K provides a complete summary of the main themes from participants' responses to these individual case studies.

Exhibit 4-2. Summary of case study examples

Case Study	Summary
Comparing Hospital Quality (hospital quality)	Used in all deliberative methods, the hospital quality case study presented evidence on the differences in clinical results between hospitals that had high or low volume for several procedures. Participants discussed whether and how evidence of differences in hospital performance should influence individuals' decisions when choosing a hospital for one of these procedures. Also, the case study asked participants to assume the role of societal decisionmakers as members of a town council who had to choose between two health plans. One plan covered services only at the high-volume regional hospital 50 miles away, and the other plan covered only services at the low-volume local hospital.
Upper Respiratory Infection in Children: Antibiotics Versus Symptom Treatment (URI)	Used in the CD and CP methods only, the URI case study focused on whether the evidence of harm from overusing antibiotics warranted societal boundaries. Participants received evidence that antibiotics, which work against bacterial infections, have no benefits for children with upper respiratory infections that are viral in nature. They were also told that an American Academy of Pediatrics guideline warns physicians not to use antibiotics unless they know the child has a bacterial infection, noting the potential for the child to become resistant to bacteria over time with overuse of antibiotics. Participants also learned about the methicillin-resistant <i>Staphylococcus aureus</i> (MRSA) infection as a societal impact of overusing antibiotics.
Obesity Management: Comparing Treatment and Prevention (obesity management)	Used only in the CP method, this case asked participants to allocate money from a Federal grant to efforts to reduce obesity. In the first location, discussion was open ended—specific options were not presented. In the remaining three locations, participants were asked to choose three of six options to reduce obesity: (1) intensive diet and exercise programs for the moderately obese, (2) bariatric surgery for the moderately obese, (3) changing the community environment, (4) changing the school environment, (5) increasing taxes on non-nutritious foods for people who buy them, and (6) increasing taxes on companies that make non-nutritious food.
Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicines (heart disease)	Used only in the CP method, the heart disease case study provided information and evidence about the relative benefits of using medicines only or stents plus medicine to treat mild heart disease. Evidence indicated that 33 percent of people with mild heart disease should receive a stent as the first line of treatment, but that 55 percent of people with mild heart disease currently receive stents as the first option. With this information, participants assumed the role of a patient advisory board to a health insurance plan to discuss advice and recommendations they should provide to the health plan.
Comparing Approaches to Preventing Illness: A Fictional Case (fictional case)	Used only in the CP method, this case asked participants to make recommendations on approaching a fictional, progressively debilitating illness called Shake, Rattle, and Roll (SRR). Participants discussed three approaches to managing SRR: (1) treating the illness with an expensive medication; (2) preventing the illness by encouraging everyone to exercise 30 minutes per day every day, which, with the costs of training everyone to exercise, was the most expensive option; (3) preventing the illness by putting an inexpensive medicine in the water supply, although 1 out of every 1,000 people who drank the water would eventually die from an allergic reaction.

Note: CD = Community Deliberation; CP = Citizens' Panel.

Findings

Here we describe the public's attitudes toward boundaries, their views on the role of evidence in decisionmaking, and concerns about health care costs. We also describe the values elicited during deliberation. The findings were often consistent across deliberative methods, but we note differences by deliberative method where readily apparent.

Although this chapter describes participants' views as they occurred throughout the discussions, we conclude this section with a brief summary of participants' responses to the deliberative question—a snapshot that captures high-level changes over the course of the deliberations.

Attitudes Toward Boundaries

When presented with the deliberative question, many participants' initial responses showed that they perceived boundaries as compulsory rules and regulations that disallowed exceptions, interfered with the doctor-patient relationship, and limited choice. Participants were hesitant to establish boundaries for health care decisions and, when discussing possible boundaries, often noted that they applied only in certain situations.

Overall Beliefs About Boundaries

No matter what type of boundary discussed, participants expressed three attitudes and beliefs related to boundaries.

Interference with autonomy. Participants believed that boundaries obstruct decisionmaking by patients and doctors, and interfere with autonomy and freedom of choice. Doctors cannot make the best decisions for their patients if their hands are tied by boundaries. For instance, one participant said, *"I firmly believe that most doctors out there honestly got into the profession because they want to help people.... And if their hands just continually get tied with more and more and more regulations, they'll be less able to do that"* (CD group in Durham, NC).

Practical challenges. Participants said that boundaries create logistical and practical challenges, and should be feasible and realistic to implement. Participants often expressed practical concerns about implementation, noting that boundaries *"just wouldn't work,"* would increase costs due to additional oversight, or would have other detrimental unintended consequences, such as stifling innovation.

Slippery slope. Participants described boundary setting as a slippery slope, saying those setting boundaries should know *"where to stop."* Participants spontaneously used the term *"slippery slope"* and shared an understanding of it: that if one limit or boundary is tolerated, it is easier for more extreme restrictions to be imposed. Participants also used other phrases, such as *"setting a dangerous precedent"* or *"opening a door to something worse,"* to denote this concern.

Views of Different Types of Boundaries

Exhibit 4-3 summarizes participants' perceptions of different types of boundaries. We grouped these boundaries into five categories, from most to least accepted by participants: (1) education, an alternative to boundary setting; (2) guidelines; (3) accountability mechanisms; (4) penalties and incentives; and (5) rules and regulations.

Education. Participants frequently proposed education for patients and providers as an alternative to setting boundaries. The majority of participants expressed the belief that education helps people make the best decision for themselves, but a minority of participants suggested that education would not always result in the right decision.

During the deliberative sessions, participants desired different kinds of educational materials for the public, including information on provider quality and cost such as discussed in the hospital quality case study, information to increase awareness of public health issues such as the overuse of antibiotics and MRSA in the URI case study, and general information on treatments and interventions to help improve their decisionmaking. Participants often spoke of doctors' responsibility to be educated themselves and to educate their patients. Some participants believed strongly from the start that patients have a responsibility to educate themselves and ask questions

of their doctors, while other participants explicitly stated that the discussion changed their perspectives about patients' role in health care. For example, during the second day of a CP group, one participant said:

Not asking the right questions to my doctor, I find, has got me just relying on them to make the decisions. And after being here I'm starting to understand that I do have a choice. And I really need to start asking questions. Young lady was sitting over here mentioned the fact that there, that they had somebody who died in their family. I can't say exactly how she worded it. But my father took, he died from leukemia and he took, we were experimenting on a new drug. And I thought anything is necessary. I never thought to myself that I could have asked these doctors how has this been studied? Any of these questions. I know now that I will never walk into a doctor's office or that anybody that I know walk into a doctor's office blind without being informed of things that they need to ask.

Guidelines. Participants generally interpreted “guidelines” as nonbinding recommendations that allowed clinicians to depart from these recommendations when appropriate. Often, the term “guidelines” was used interchangeably with “best practices” or “standards.” Participants identified doctors, experts with specialized education, as best equipped to make decisions for individual patients based on evidence. Thus participants wanted clinicians to be aware of and generally follow the guidelines from medical professional societies, but they also wanted clinicians to be allowed to depart from the guidelines when needed for individual situations.

Participants' perceptions about guidelines may have varied depending on exposure to multiple case studies. Participants in BCD and ODP, the two methods with only the hospital quality case study, frequently described existing guidelines as effective, while participants in CD and CP, who discussed the URI case study, talked about how guidelines could be ineffective. Analysts observed these differences by method when conducting the thematic analysis; however, these observations have not been confirmed with an indepth analysis comparing responses by deliberative method.

Accountability mechanisms. Participants often called for increased “accountability” of doctors—that is, oversight to hold them responsible for their actions. Participants discussed accountability mechanisms in the context of both guidelines and rules and regulations. Mechanisms of accountability included having doctors verify test results of a bacterial infection before prescribing antibiotics, tracking numbers of prescriptions, or monitoring and reporting on doctors or hospitals. Although some participants voiced concern that this type of oversight would constrain doctors' ability to tailor care for their patients, others saw this oversight as a way to enforce the guidelines while still being flexible. Also, some participants connected accountability with concerns for the community, which included the belief that doctors have a responsibility to protect the community from harm.

Penalties and incentives. Participants often used the terms “incentive” or “penalty” interchangeably. A higher insurance premium paid by a smoker, for example, might be viewed as a penalty for choosing to smoke or as an incentive to stop smoking. Examples of penalties and incentives included financial (e.g., taxes, fines, nonreimbursement for services, discounts); professional rewards or sanctions (e.g., loss of license, accreditation, ratings, quality awards); or legal consequences (e.g., jail for serious harm). Participants were divided on their support for penalties and incentives for doctors and patients. Some supported penalties and incentives they

believed preserved choice and encouraged better decisionmaking, such as professional rewards or discounts for going to a high-volume hospital. Others expressed concern that reimbursement incentives could negatively affect clinicians' integrity by prioritizing financial gain over the patient's health, or that they may be unfair to some people, saying "*what's a good decision for one person may be different for someone else*" (ODP group, remote). This topic was discussed to a lesser extent than other boundary topics.

Rules and regulations. Participants interpreted "rules and regulations" as mandates or "placing limits" on what doctors can recommend or prescribe. Participants were not initially in favor of restricting clinicians in this way, as many participants believed that doctors departed from the recommended treatment only due to specific knowledge of the patient.

When talking about rules and regulations, some participants cited parallels with existing public health laws that they did accept, including regulation of controlled substances, monitoring of pseudoephedrine purchases, and traffic laws such as speed limits, traffic lights, and seatbelts. Similarly, participants referred to existing limitations on health care choice, such as health plan network restrictions and, to a lesser extent, health plan treatment restrictions. Participants sometimes acknowledged the necessity of these choice restrictions, but they disliked them both in principle and in practice, especially in cases when distance from the nearest hospital hinders access to care. As the sessions progressed and participants discussed the issues currently facing the health care system, some participants, especially in the CD and CP methods, made choices in the case study exercises indicating that they were more accepting of regulations and rules on doctors. Yet participants were less accepting of "placing limits" on patients' choice of doctor or hospital, despite acknowledging that limitations from insurance plans on in- and out-of-network providers already exist.

Exhibit 4-3. Participant perceptions of boundaries

Approach	What is it?	Who is it for?	What is the purpose?	Who implements?	What are perceptions?	Implications/ Will the boundary work?
Education and boundaries that provide or require education	Education for patients included information about health issues, treatment options, quality, and cost. Doctors expected to and responsible for educating themselves on current research.	Patients, doctors	Encourage better decision-making	Medical associations, government	Strongly positive. Helps people make the best decision for themselves. Perceived doctors as best equipped to understand and interpret evidence and inform patients, but patients have responsibility to be educated.	Perceived that education would improve decisionmaking. Limited discussion about how education may not always result in the “right” decision.
Guidelines/ standards	Nonbinding recommendations such as best practices or standards.	Doctors, hospitals	Ensure quality of care	Medical associations	Positive. Preserves doctor autonomy, allows exceptions. Acknowledged that providers do not follow guidelines consistently.	Initially perceived that guidelines would work. Concerns that guidelines might not work triggered accountability discussions. Noted difficulties with enforcement.
Accountability mechanisms	Oversight of and adherence to guidelines or rules/ regulations.	Doctors, hospitals	Ensure quality of care and protect patients from harm	Medical associations that accredit or license doctors or hospitals, government	Positive. Seen as an alternative to rules and regulations and as a way to enforce guidelines, while still being flexible.	Some believed it would work; others did not. Perceived as enforcement of guidelines, but may constrain doctors’ ability to tailor care.
Penalties/ incentives^a	Used terms interchangeably. Examples included financial; professional rewards or sanctions; legal consequences.	Patients, doctors, hospitals	Encourage better decisionmaking	Medical associations, insurers, government, other payers such as employers	Varied, depending on underlying boundary. Seen as a way to enforce accountability mechanisms or rules and regulations.	Could improve decisionmaking, but financial incentives could compromise doctor’s ethical practice.
Rules/ regulations	Mandates or restrictions; no exceptions on what doctors can do and patients can choose.	Patients, doctors, hospitals	Protect patients from harm or ensure fair allocation of resources	Insurers, government	Negative. Did not like, but acknowledged necessity in certain situations. More accepting of regulations on doctors than patients. Discussions about harms to patients and community; included physical and economic harm.	Perceived as effective, but with unintended consequences limiting access to care and a slippery slope to erosion of choice and compromising physician judgment. Also raised practical challenges with implementation.

^aLimited discussion on this topic, primarily in response to probes in the hospital quality and obesity management case studies.

Acceptability of Boundaries

Overall, acceptable boundaries were those that encouraged or protected what is in the best interest of the individual patient. Participants rarely expressed discomfort with boundaries that clearly and directly benefited patients. For example, even though government intervention in health care was perceived negatively, government mandates requiring transparency and providing information about hospital cost and quality were typically viewed positively. Participants thought this type of mandate encouraged informed decisionmaking and preserved choice. Similarly, educating doctors about what works best through guidelines and providing professional oversight were also typically viewed more positively; they were perceived as ensuring standard levels of quality for patients.

Participants displayed more disagreement and discomfort when boundaries caused conflict between what is perceived as being in the best interest of individual patients and what is in the best interest of community or society. In these instances, participants responded more positively toward boundaries that allow for exceptions and preserve autonomy for the doctor and choice for the patient. Often, during deliberation, participants tried to identify the “middle ground” between guidelines with little oversight versus rules and regulations with no exceptions.

Perceptions of Entities Setting Boundaries: What Drove Participants’ Trust?

Participants immediately noted that health insurers or the government typically set boundaries in health care. Further discussion elicited other entities, such as medical associations, employers, the general public, or all groups together. Overall, participants trusted independent medical associations more than insurers, the government, or employers.

Three main elements appeared to influence trust of these different entities: whether boundary setting is an appropriate role, perceived expertise, and perceived motivation. Exhibit 4-4 summarizes participant perceptions of these elements for the four main boundary-setters—medical associations, government, health insurers, and employers.

Appropriate role. Participants perceived medical associations as clearly having a role in boundary setting by educating physicians through guidelines based on the latest and best medical evidence. Almost all participants were familiar with insurers’ roles in limiting care and accepted that a component of the insurers’ role was to limit care. Participants debated whether other payers, such as the government or employers, have the “right” to set some boundaries. Some described health care coverage as an entitlement and opposed the idea of payers setting boundaries. Other participants believed that payers have the right to set some boundaries based on quality and cost because they pay the majority of costs. For example, participants did not initially perceive a role for government involvement, but upon learning that the government is the largest payer, some participants were more willing to accept government involvement in health care cost containment. Similarly, although in the minority, some participants stated that employers have the right to make health choices for their employees because employers offer the plans as a benefit to employees, who are free to choose another employer.

Exhibit 4-4. How do participants perceive boundary-setters?

Entity	Is boundary-setting an appropriate role?	What is the perceived expertise?	What is the perceived motivation?
Medical association	Yes, as independent and objective groups to guide physicians.	Medical expertise. Individual physicians understand needs of their patients.	Primary motivation is concern for patient health and welfare. Not motivated by money.
Health insurers	Yes, as they pay for the health care for their members.	Lack of medical expertise; do not know individual patient.	Primary motivation is profit, as a business, often at expense of quality. Minority view thought insurers could help patients get quality care.
Government	Not initially. Primary duty is not to interfere with patient's autonomy. During deliberation, other roles emerged, such as the duty to protect citizens from harm and pay for health care.	Lack of medical expertise, with exception of a few agencies, such as the Centers for Disease Control and Prevention or U.S. Food and Drug Administration. Do not know individual patient.	Divided. Some perceived as well intentioned, while others saw government as corrupt or ineffective.
Employers	Divided. Employers pay for health care as part of compensation package. But employer should not be involved in issues related to personal health care or health status.	Lack of medical expertise. For large employers, do not know individual patient.	Divided. Invested in employees and want them to be productive, but also need to contain costs to make a profit.

Motivation and expertise. Participants perceived medical associations as independent, with no financial stake in care practices or decisions, and as having the needed medical expertise. For the other entities, participants had negative or divided perceptions of motivations for and expertise in setting health care boundaries. For government involvement, some participants indicated that government was well intentioned but ineffective; others described government as corrupt and ineffective. For insurers and employers, many participants expressed concern that most players involved in the business of health care make decisions aimed to increase profit at the expense of quality.

Thus, while participants could accept boundaries limiting choice of hospitals based on evidence of better clinical outcomes in the discussions, they were skeptical that, in reality, insurers would establish boundaries on anything other than cost. A notable exception was discussion related to a large insurer, primarily in Sacramento, CA, and ODP groups; participants often acknowledged that this insurer constrains choice of provider or treatment to manage costs, but also appears concerned with quality: *“You may be in Kaiser, but yet Kaiser sends you to Mercy, because Mercy is renowned—known for their heart catheterization care. So they contract, because that’s where they do it the best. That’s what I’ve experienced here, in my town”* (ODP group, remote).

Some participants saw the appeal of multiple groups working together to set boundaries, acknowledging that different groups brought different interests. For example, one participant said, *“A team of doctors, insurance companies, coming together. Not that just everyone would sit down and get to make the decisions, but people that are informed, we get to make the rules that are agreed upon by the doctors and the insurance companies”* (BCD group in Sacramento, CA).

A minority of participants raised the need for patient input in any guidelines or regulations. Most of these participants were primarily concerned that without patient involvement, there would be minimal patient buy-in and understanding. Some participants mentioned that they wanted to make sure that patient priorities were included.

Role of Evidence in Decisionmaking

In this section, we present participants' overall views of evidence, describe how participants applied evidence of medical effectiveness and evidence of physical and economic harms, and summarize how participants applied evidence to individual and societal decisions.

Overall Views of Evidence

How participants defined evidence. Participants rarely focused on the concept of evidence when initially discussing the overarching question and rarely questioned the evidence presented in the case study. Rather, over the course of deliberation, participants discussed the relative importance of different types of evidence and the role evidence plays in decisionmaking, described below in more detail.

Participants often discussed evidence using the specific terms presented in the case materials. For example, participants used the terms “*success rates*” or “*clinical results*” as evidence in the hospital quality case study, or referred to diagnostic and screening “*test results*” as evidence in the URI case study.

Aside from the case studies, other comments indicated that participants equated evidence with experience—the doctor’s accumulated experience and clinical judgment, personal lived experiences, or common sense (e.g., “*evidence that the sun’s coming up tomorrow*”).

Usefulness of evidence. In general, participants viewed evidence positively and stated they valued it highly in making informed health care decisions. Patients wanted more transparent information about quality or clinical effectiveness and believed that patient access to information about evidence is limited. For example, one participant stressed, “*The one thing I would say adamantly is that the information must be provided to each and every patient so they can make an informed choice [when choosing a hospital]*” (ODP group, remote).

Yet the emergence of two beliefs about evidence highlights participants’ views on the limitations of exclusively using evidence in decisionmaking. First, evidence of what works for most people may not apply to each patient, as “*everyone is different.*” Participants frequently mentioned that what works for each individual may vary and expressed concern about evidence-based boundaries. These participants wanted to avoid the “*one size fits all*” or “*cookie-cutter medical treatment*” model, and stressed the importance of the doctor’s personal experience with the patient that might outweigh or contradict the evidence. Participants often used terms such as “*exception*” or “*outlier,*” or allowing doctors to be able to go “*outside the box.*”

Second, evidence was viewed as imperfect: changing over time, often excluding specific populations based on age or ethnicity, lacking clarity, and at times nonexistent. Some participants discussed how medical conditions can differ for certain populations and questioned whether the medical studies accounted for differences in age or ethnicity. Other times, participants questioned the number of patients in the study or asked if hospital ratings considered population differences. Some participants questioned whether evidence is always up to date or acknowledged that evidence may not yet exist for treatments.

Relative priority of evidence versus other factors. A variety of other considerations competed with evidence in health decisions, including—

- Personal preferences, medical histories, and unique situations
- The doctor’s clinical judgment
- Cost concerns, such as individual out-of-pocket expenses

Often, these other factors became more important when participants did not see the relevance of the evidence to the situation or if the evidence they wanted did not exist.

Participants also assessed the relative priority of clinical outcomes versus patient-centeredness. When discussing the first part of the hospital quality case study in the BCD, CD, and ODP methods—which focused on individual decisionmaking with respect to choice of hospital—participants prioritized evidence of clinical outcomes, such as survival, and treating patients with respect. Many participants argued that clinical outcome is the most important factor in quality health care and wanted to get the “*best care*,” no matter the cost or distance one had to travel or whether the facility or staff were “*nice*.” Others prioritized receiving personal and respectful care, in which the patient is more than just a “*number*” at a high-volume hospital. In addition, some of these participants believed that caring and respectful treatment has an impact on clinical outcomes for individual patients; for example, one participant said: “*There's a lot more to health care than just the technical aspects of it. Did they treat you well? Were you comfortable? Were they responsive to your needs? All those things, I think, make a big difference in the patient's welfare*” (CD group in Silver Spring, MD). Others argued that a good-quality hospital would both produce good clinical outcomes and treat patients with respect, negating the need to choose one at the expense of the other.

Doctors’ role in applying and communicating about evidence. Participants identified doctors, as experts with specialized education, as best equipped to evaluate whether to comply or depart from the evidence for the best interests of the individual patient. Further, participants often identified the doctor as responsible for discussing the evidence of benefits and harms with patients so that patients can make informed decisions. One participant said: “*It's up to the doctor to inform [patients] completely. Not just say from their perspective what they think is the best thing to do...but you also want to tell people the whole story. They don't want a half truth. You want all of it*” (CP group in Chicago, IL).

Applying Evidence of Medical Effectiveness

Perceptions of medical effectiveness. Participants referred to medical effectiveness as whether a treatment or surgery “*works*” or “*works well for most people*.” Occasionally, participants said that effectiveness is based on the “*average*” of what works for most people.

In considering societal limits, evidence of medical effectiveness alone was not compelling enough to shift attitudes toward stricter boundaries unless cost or harm was also a factor (as described below in applying evidence of physical and economic harm). But even if not explicitly stated, participants appeared to assume that boundaries would be based on evidence of medical effectiveness.

Exhibit 4-5 summarizes participants’ perceptions of when different levels of evidence of medical effectiveness can be used in decisionmaking, as well as other factors that are considered. Although participants did not explicitly describe the different levels of evidence noted in the

table, their comments indicated how they perceived application of evidence from these different levels.

Exhibit 4-5. How participants perceived different levels of evidence of medical effectiveness

Type of Evidence	Usefulness in Personal Decisionmaking	Usefulness in Boundary Setting	Exceptions to Using This Type of Evidence	Other Factors That Are Considered
Clear evidence of medical effectiveness	Highly useful when considering a personal medical decision, such as choosing treatment/doctor	Highly useful for encouraging quality care through guidelines, but not sufficient for prohibitive restrictions on care	<ul style="list-style-type: none"> Evidence seen as “not applying to me” Evidence is for a medical condition that seems rare Patients have specific religious or philosophical beliefs (minority view) 	<ul style="list-style-type: none"> Personal preferences for treatments or providers Lived experience of person or family with specific treatments or providers
Clear evidence of medical ineffectiveness	Highly useful when considering a personal medical decision, such as selecting a treatment	Moderately to highly useful for setting boundaries such as insurance coverage limits, but not sufficient for prohibitive restrictions on care	A patient is dying, has tried all other options without success, or wants to pay for it out of pocket	<ul style="list-style-type: none"> Not discussed by participants
Uncertain evidence of medical effectiveness	Somewhat useful when considering a personal medical decision, such as selecting a treatment	Limited usefulness in boundary setting, as it is difficult to apply to individual situations	<ul style="list-style-type: none"> Evidence is seen as “not applying to me” A patient is dying, has tried all other options without success, or wants to pay for it out of pocket 	<ul style="list-style-type: none"> Personal preferences for treatments or providers Costs to individuals or communities Absolute risk and personal comfort with risk
No evidence of any kind^a	Not considered useful	Highly useful for setting limits on coverage, as society is not obligated to pay for treatments with no evidence	A patient is dying, has tried all other options without success, or wants to pay for it out of pocket	<ul style="list-style-type: none"> Evidence about effective treatments Personal comfort with risk

^aLimited discussion on this topic.

Clear evidence of effectiveness. Participants appeared to want clear and compelling evidence as a prerequisite to limiting choice or autonomy. For example, one participant described convincing evidence as *“This has worked on 99.5 percent of the people who prescribed it to. It works. It tells me, and we did the study over a period of 8 years, and [had] so many—you know, 50,000—people”* (BCD group in Chicago, IL).

Clear evidence of ineffectiveness. If evidence clearly shows a treatment to be ineffective, participants were generally comfortable with setting some restrictions. For instance, one

participant explained: *“I think that a patient should be able to choose for it but I don’t necessarily think that insurance or the government should pay for it if it’s shown to [have] zero effectiveness”* (CD group in Durham, NC). Also, participants who understood that antibiotics have no effect on viruses were comfortable restricting their use in patients known to have viruses. A few also discussed the possibility of requiring patients to use the evidence-based treatments before trying the less effective treatments that might work well for them.

Uncertain evidence. Participants struggled to justify restricting care when evidence is uncertain. Participants described uncertain evidence as more than just situations where the research results are mixed or the evidence itself is unclear. Many comments reflected participants’ perspective that evidence is uncertain when it may not apply to specific patients in specific situations. For example, one participant said: *“Acupuncture, most of the time, is not covered by health insurance, but some people swear by it. So I would hate to see, ever, that a doctor’s hands were tied by the government as to what kind of treatments are available”* (BCD group in Chicago, IL). Further, participants discussed situations when the *applicability* of evidence was uncertain, such as the use of antibiotics in a patient who may or may not have a viral complaint or whose personal medical situation may affect the applicability of the evidence.

No evidence at all. When there is no evidence at all (either of effectiveness or ineffectiveness), participants did not obligate society to pay for those treatment options, but said that individuals could choose to pay for those options themselves. But participants still had concerns about limited choice: *“Not every potential treatment has research and, especially in our society in the U.S., we tend to focus on traditional Western medicine...If we’re limiting ourselves to our framework that could discount something that could work for someone that just doesn’t fall into that evidence-based practice...”* (CD group in Sacramento, CA).

Applying Evidence of Physical and Economic Harm

Perceptions of harm. Initially, most discussions focused on risks of direct physical harm to patients or families. Often, participants described physical harm as pain, increased risk of future injury or illness, or death. Some comments, however, also included emotional and psychological harm (e.g., anxiety about expected outcomes). Participants frequently considered potential harm as it would affect someone they love—for example, a child or parent. Participants responded more viscerally to the prospect of harm to children and admitted making emotional decisions to help their child feel better.

Participants weighed the evidence of medical effectiveness along with the evidence of harm. In weighing these risks and benefits, participants often identified tradeoffs in individual decisions, such as weighing the benefits and risks of invasive surgery in choosing among treatment options or weighing side effects and benefits of taking a medicine. For example, one participant said: *“The only time I ever question a physician is when I find out what the side effects are. I always ask, ‘What will this do to me?’ And if the side effects outweigh what the medicine will do, then I’ll either ask for different medication or preferably not take it if I disagree”* (CD group in Durham, NC).

Later in the deliberative sessions, participants discussed harm that resulted to others in the community. Harm to others could be physical or economic. Most of the time, discussions of broader harm did not arise spontaneously but arose from additional information in the case studies. For example, in the URI case study, overuse of antibiotics resulted in physical harm to others from the emergence of “super bugs” such as MRSA. In the hospital quality case study,

economic harm to the community resulted from the closure of the low-volume local hospital if participants, as the town council, chose a certain health plan.

Across all discussions of harm, participants' perceptions of risk varied, as did the level of comfort with risk-taking. For example, in examining the differences in risks of complications between the low-volume and high-volume hospitals, some perceived a high level of risk at the low-volume hospital, while others did not perceive much difference. These different perceptions could have been due to participants' difficulty in understanding the statistics presented, or the differences could reflect, as some participants pointed out, that each individual's comfort with risk of harms could be different. When considering societal limits, these relative assessments of risks often influenced attitudes about whether society should establish a boundary: the greater the danger, the more support for the boundary.

Exhibit 4-6 summarizes participants' perceptions of when evidence of physical harm and economic harm are useful in decisionmaking, as well as other factors that are considered.

Exhibit 4-6. How participants perceived different types of evidence of harm

Type of Evidence	Usefulness in Personal Decisionmaking	Usefulness in Boundary Setting	Exceptions to Using This Type of Evidence	Other Factors That Are Considered
Evidence of physical harm	Highly useful when considering a personal medical decision, such as selecting a treatment or doctor	Highly useful for setting restrictions on care. Clear evidence and severe harm to individuals or communities are the only acceptable basis for prohibitive boundaries.	<ul style="list-style-type: none"> • Risk of physical harm occurring is low or physical harm is minor • Harm accrues to few individuals 	<ul style="list-style-type: none"> • Personal costs • Personal preferences for care
Evidence of economic harm to community or society	Rarely discussed in the context of personal health decisions	Moderately to highly useful for setting boundaries such as coverage decisions, but depends on scale of economic harm to community or society. Particularly relevant when costs are high.	<ul style="list-style-type: none"> • Costs are low • Economic harm accrues to few individuals • Implementing a boundary results in higher risk of physical harm 	<ul style="list-style-type: none"> • Personal preferences for care • Evidence of effectiveness or quality of care

Evidence of physical harm. Most discussions of physical harm occurred around the URI case study. Of those who discussed that study, many accepted that society should reduce overuse of antibiotics to prevent long-term harm to the community from MRSA and ensure that antibiotics remain effective for future generations. But some struggled to name an acceptable level of risk, especially when the boundary limited choice. For example, one participant compared restricting use of antibiotics to eliminating peanut products in schools, saying: *“Because one child has a severe peanut allergy, nobody's allowed to bring a peanut butter sandwich to school. I want to protect all the children, but I don't know if I want to protect this one child at the loss to my own. So it's a lot harder when you slide the scales down”* (CD group in Durham, NC).

In most discussions of the URI case study, the preferred boundary to protect others from harm consisted of guidelines and oversight by medical authorities. Often, evidence of harm was a compelling factor in shifting attitudes toward stricter boundaries. For example, one participant in a CD group in Silver Spring, MD, said: *“I change my mind. I really—I'm still very resistant to it...Because [at the] last meeting I want to have the freedom to put my foot down and say I want*

this antibiotic. But that 5,500 deaths of MRSA has just been rolling around in my head and maybe—it's hard because I don't like the idea of regulation and things like that. Although you're right, the vaccinations and things. I was vaccinated for polio and I think everyone should." In other instances, participants cited stricter rules to prevent adverse effects on the health of the public, such as requiring people with tuberculosis to receive treatment. Below is a deliberative exchange from a CD group in Sacramento, CA, about possible stricter rules:

Male 1: They're trying to come up with how do they get the doctors to quit prescribing it unnecessarily.

Female 1: They should put them on a probation or something, and so it's listed –

Female 2: Take their license away.

Female 1: You know, like, health, when you go to restaurants if you have a green [sign] in your window.

Female 2: Like a report card, yeah.

Female 1: Well, there are report cards with the Department of Managed Health Care.

Male 1: Yeah, there should be some reprimands.

Female 3: When we do our taxes sometimes, we can just be randomly audited. So if they had it to where doctors could realize that they could be randomly audited or checked up or investigated, just them individually, then maybe they would make sure that they dotted all their—crossed their t's and dotted their i's and made sure they were doing things right.

Evidence of economic harm. Discussions of economic harm arose most often in the context of the hospital quality case study, in which, when participants were asked to take on the role of town councilors, they weighed the clinical quality for those who needed specialized surgery against the economic impact and loss of local access to care for the community. Most participants cited one of these harms in providing a rationale for their selection of one of the boundaries on choice offered in the case study. Many participants stated that the economic impact and loss of local access to care to the whole community outweighed clinical quality for those few who needed specialized surgery. One participant explained: *"If they send everybody to this high-volume hospital and they lay off half the workforce of their town, that's going to have repercussions on their bottom line....I think they would have to think about how this would financially impact the town as a whole"* (ODP group, remote).

Of note, a vocal minority argued for a way to improve the local hospital and eliminate the need to limit choice. These participants generally wanted to improve quality at the local low-volume hospital so that patients get the highest quality care at the local hospital to prevent it from closing.

Also in relation to economic harm, participants discussing insurance as a pooled set of shared resources nearly unanimously supported the need for boundaries to prevent people from overusing their share of the resources. For example, one participant said: *"I think when there's a cost to society, I think it's well within society's mandate to take a tape and create boundaries"* (BCD group in Chicago, IL). Most participants believed that people will try to take advantage of the system if some boundaries are not in place. For example, a participant in CP said: *"We cannot offer everything to everybodyAnd since we cannot reverse the fact that all of us are paying for the health care of others in one way or another, we have to collectively come to some*

compassionate limits because some people will abuse the system” (CP group in Durham, NC). However, participants did not usually discuss where these limits should be or how changes should be achieved. Despite the predominant view that society’s resources need to be safeguarded, some participants, across methods and groups, supported providing a level of basic coverage to all consumers. But it was unclear how comprehensive participants intended this coverage to be. For example, one participant stated: “I think there should be some minimum level of health care for everybody and what that minimum level is, I don’t know. It’s probably what we can afford would be the determining factor. But we can’t do everything for everybody, particularly people who are in bad health because of lifestyle choices they make” (ODP group, remote).

Applying Evidence to Personal and Societal Decisions

These discussions so far, and Exhibits 4-5 and 4-6, show how participants judged the usefulness of different types of evidence. Below, we summarize how, during the case studies and exercises in the deliberative sessions, participants said they would apply this evidence in making treatment decisions, selecting providers, and setting limits on care for their community or society.

Making treatment decisions. In considering personal medical decisions, participants’ comments indicated they placed high value on evidence of medical effectiveness and evidence of physical harm, and a lower value on evidence of economic harm. Participants placed similar value on using both certain and uncertain evidence, although evidence of any kind might quickly be discounted if it was seen as “not applicable to me.” Participants indicated that they might consider treatments for which no evidence existed in certain circumstances, such as participating in drug trials or trying untested treatments as a last resort.

Selecting a provider or hospital. Although many participants wanted to select the hospital with the better clinical outcomes, some participants placed a lower value on evidence of clinical effectiveness for this decision. These participants said they were not willing to have this evidence drive their decision, particularly when they considered that the better quality provider or facility might have significant drawbacks to them personally. For example, some participants found reasons to devalue or ignore quality measures for hospitals in favor of going to a hospital they considered more convenient or where they might receive more respectful treatment.

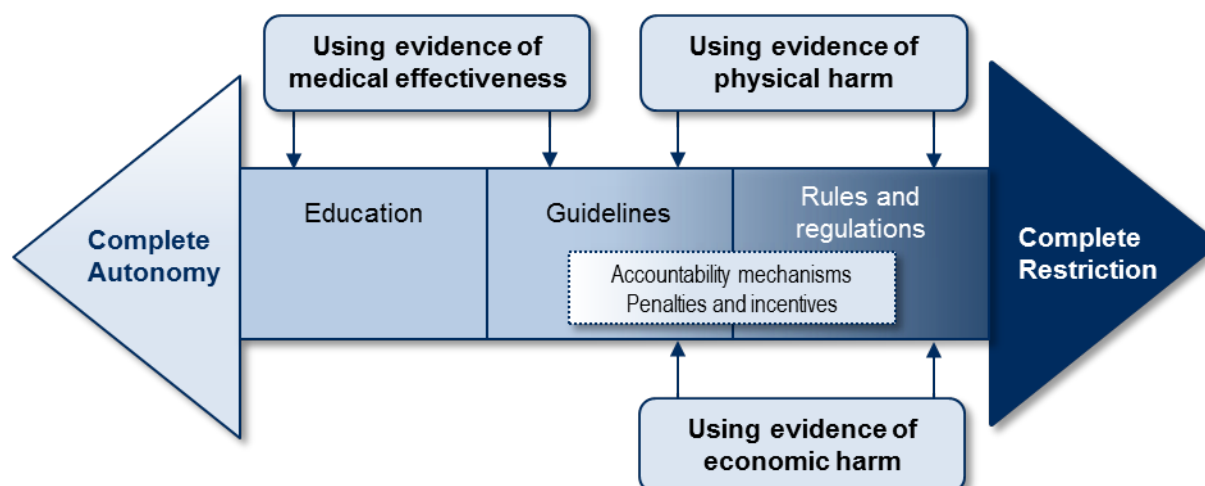
Setting limits on care for the community or society. When making decisions or setting limits for a community or society at large, participants placed a high value on evidence of physical harm, evidence of economic harm, and clear evidence of medical effectiveness/ineffectiveness. Participants placed a lower value on uncertain evidence of medical effectiveness. An absence of evidence of medical effectiveness was not discussed in the context of societal decisionmaking.

Notably, evidence of physical harm was the only acceptable basis for prohibitive restrictions on care, but evidence of economic harm to the community also had a great impact. Many participants sacrificed better clinical outcomes for a few on the basis of uncertain evidence that the local hospital would close. Others were willing to sacrifice clinical outcomes only with clear evidence that economic harm would result to the community.

In Exhibit 4-7, we show how participants’ perspectives on using different types of evidence could be related to the different approaches for boundary setting, ranging from complete autonomy (no boundaries) to complete restriction (strict boundaries). Evidence of medical effectiveness alone was not compelling enough to move toward stricter boundaries, but was

valued in individual decisionmaking and encouraging better quality care through guidelines. Consideration of evidence of economic harm was a more compelling factor than medical effectiveness in limiting choice, and evidence of physical harm was the most compelling in establishing stricter boundaries.

Exhibit 4-7. Using evidence as a basis for boundary-setting approaches



Cost Considerations for Setting Boundaries

Even though participants discussed how evidence of medical effectiveness, physical harm, and economic harm could be used in boundary setting, many participants expressed concern that the primary motivation in establishing boundaries was cost containment rather than ensuring access or quality. This concern manifested in discussions of how costs limit access to services; the need for more transparent cost information; relationships among boundaries, cost, and quality; and cost implications for the larger society.

Limiting access to services. Participants noted that costs already constrain patients' choices of and access to certain services. Participants discussed personal costs of services ranging from basic health care to specialized care and elective surgeries, including out-of-pocket expenses, health insurance premiums, and travel expenses (i.e., 50 miles to the high-volume hospital for surgery). Often, participants expressed the belief that *"more money equates to more choices"* and noted that ability to pay can expand access to out-of-network providers, as well as treatments not supported by evidence or "experimental" treatments that insurance will not cover. Other comments suggested that someone's ability to pay should not determine whether a patient receives needed care or the quality of care a person receives.

Need for transparent cost information. In general, participants noted the difficulties in figuring out how much health care costs; they expressed that more transparency of health care costs would benefit the public. Multiple participants shared personal experiences of not being able to predetermine costs or referred to *"waiting for the bill to arrive,"* while only a couple of participants referred to success in finding out the costs of treatment ahead of time. Although only a small number of participants explicitly mentioned the free market, many others indirectly acknowledged that health care does not function as a free market because people do not know,

and cannot ascertain, what treatments cost in advance. Because of these challenges, participants viewed government mandates requiring transparency about hospital cost and quality positively.

Cost and quality. Participants discussed the relationships among costs, boundaries limiting choice of provider, and quality of health care or health insurance. Many comments indicated the belief that better care is more expensive, and boundaries aimed at cost containment limit access to that better care. For example, some participants found it counterintuitive that a high-volume, higher quality hospital could be less expensive than a low-volume, lower quality hospital. Fewer comments overall indicated that better care does not need to be more expensive, identifying less invasive procedures or the complications associated with poor-quality surgery. Also, while discussing the hospital quality case study, many participants said that insurance plans with more provider choices were better and thus more expensive.

Cost implications for society. In addition to discussing what health care costs to them, participants explored how society pays for health care and who is responsible for paying these costs. Many participants noted that we all pay collectively for health care through taxes or insurance premiums and that there are limited resources. As described earlier, participants discussed, and were divided about, whether payers have the right to set boundaries if they pay the majority of the costs. Participants were also divided on whether or not they would be willing to pay more for everyone to have more choice. For example, when one facilitator asked participants to consider increasing premiums to accommodate more hospital choices, some participants were comfortable with an increase of up to \$25. Other participants did not want their premiums to rise, particularly if it allowed others to make “poor” choices, such as selecting less effective care, making poor lifestyle choices, or choosing to live far from a hospital.

Further, participants discussed the high costs of health care as a problem for society, with some relating high health care costs to the country’s deficit. Participants identified various factors that drive up health care costs, including profit motivations by doctors, insurers, hospitals, and pharmaceutical companies; patients who abuse the system or who do not have insurance (e.g., “*people who don’t have insurance go to the emergency room for colds*”); litigation (e.g., “*if a doctor doesn’t do certain things and then there’s not a good outcome, they get sued*”); and high costs for end-of-life care and diseases resulting from lifestyle choices. Of note, some participants argued that stricter boundaries would exacerbate these drivers by prioritizing financial incentives and profit motivations over the doctor’s clinical judgment of what is best for the patient.

Some participants discussed how personal choices affect the cost of health care overall and noted the importance of having a conversation about cost control. Some of these participants said that additional choices drive up the cost of care: “*The idea of total choice works, but when you start adding cost to it, that creates significant problems with the whole thing, because most people are confronted with the fact that they cannot afford health care at that level*” (CD group in Silver Spring, MD).

Values Exhibited During Deliberation

Values are deep-seated core convictions that individuals hold about what is most important in the decisions affecting their lives and the world around them. They are often established early in life, typically influenced by family, culture, religion, and/or country, and they tend to be enduring.¹⁹ Ethics are the moral principles that govern individual or societal actions. In the field of health

care, ethical principles typically reference how society expects medical professionals and other social institutions, such as government and business, to function.

Participants infrequently referred to the terms *values* or *ethics* in general conversation; the research team inferred them by what participants identified as important and gave as their reasons for their choices in the deliberations about boundary setting.

Exhibit 4-8 lists the main values and ethics derived from the deliberation and how participants expressed those values. Tensions between individual freedom and the greater good were most often and explicitly discussed. Other values emerged during the discussions but were less dominant—that is, they were stated less explicitly and discussed to a lesser extent. Because individual freedom and choice was the primary value expressed, we examine how other values emerged to complement or conflict with this value.

Exhibit 4-8. What kinds of values were exhibited during deliberation?

Value	How Participants Exhibited Value During Deliberation
Primary Values	
Individual freedom: Focus on personal choice and freedom from restraint	<ul style="list-style-type: none"> • Often explicitly stated by participants as a value • Reluctance to establish boundaries on individual decisions; objections to absolute boundaries • Skepticism of boundary-setters' motivations • Reluctance to apply evidence to all, as everyone is different • Belief that insurance plans with more provider choices are higher quality and willingness to pay more for insurance plans with more provider choices
Greater good: Concern for the good of the community	<ul style="list-style-type: none"> • Willingness to set boundaries on patient and physician autonomy when evidence of physical harm to others (e.g., overuse of antibiotics) • Prioritizing the economic impact and loss of local access to care for the whole community over clinical quality for those few who needed specialized surgery • Prioritizing community interventions over bariatric surgery in the obesity management case study (CP only) • Prioritizing quality of life for most people in the community in the fictional case study (CP only)
Other Values	
Fairness: Treating people equally and without bias	<ul style="list-style-type: none"> • Nearly unanimous support for boundaries to prevent people from overusing their fair share of resources, when discussing insurance as a pooled set of shared resources • Debate about whether the payer has the right to set boundaries • Concern for those who cannot receive care based on the ability to pay • Positive view of permeable boundaries allowing for exceptional situations
Personal responsibility: Being accountable for individual actions	<ul style="list-style-type: none"> • Strong preference for education and information as alternative to boundaries • Personal desires for education and information about cost and quality • Identifying patient responsibilities, such as maintaining a healthy lifestyle or getting preventive care
Respect for science and innovation: Faith in science and its tools	<ul style="list-style-type: none"> • Descriptions of evidence of effectiveness as a primary factor in personal decisionmaking • Assumptions that boundaries would be based on evidence of medical effectiveness • Need for clear and compelling evidence to limit choice or autonomy • Concern that boundaries could stifle innovation • Faith in medical associations
Pragmatism: Active interest in workability and practicality	<ul style="list-style-type: none"> • Expression of practical concerns about implementing boundaries, such as having the ability to monitor or enforce, or identifying unintended costs or negative consequences of boundaries

Note: CP = Citizens' Panel.

Individual Freedom

Participants expressed the value of individual freedom by consistently focusing on the primacy of personal choice and reacting negatively to types of boundary setting that restrict, rather than support, choice. Often, participants explicitly stated this view as a value. For instance—

It comes down to my libertarian values. I strongly believe in personal liberty. I don't want other people telling me what I can do... There's one element in the world we can control directly, our bodies, and we should be able to. I'm not comfortable with somebody else coming into my life and telling me how I can treat my body... and I'm not comfortable doing that for other people (BCD group in Sacramento, CA).

The right to make one's own decision extended to the right of parents to decide for their children, even with the possibility of future harm to the child. For example, some participants indicated they wanted the ability to decide with their doctor whether antibiotics would be prescribed, even with the possibility that bacteria would become resistant to a particular antibiotic that would not work for their child in the future.

Greater Good

Concern for the welfare of the “community at large” was a focus when participants discussed the notion that unchecked individual freedom might harm others physically or financially. Since attention to the common good usually meant a diminishing of individual freedom, these two values conflicted and were the primary tension that participants dealt with during the deliberative sessions. One participant said—

None of us exist in a vacuum—we're all part of this community, so, to as much an extent as possible, considering the other members of your community in terms of the impact of your actions is really important. I mean that's one of my values and... If you're only considering yourself and not the rest of the community, what if everyone acts that way... well, if we all do that, then we're all [doomed] (CD group in Sacramento, CA).

When placed in an “official” position as social decisionmakers, such as the town council in the fictional case, participants typically accepted the role of overseers of the common good. Some acknowledged that their priorities were different when they considered what was best for them personally versus what was best for the community. Interestingly, participants tended to believe that other members of society were less likely to care about evidence of effectiveness than they were to care about job loss. For example, one participant explained: “*I'm the town council now. We're not just thinking about individuals. Losing jobs is the most important thing to try to avoid*” (CD group in Silver Spring, MD).

Fairness

Participants most often referenced the value of fairness when discussing costs associated with the health care system. Fairness was viewed from a number of perspectives, including what is right and just and what is reasonable, given who is paying.

Fairness emerged when participants discussed allocating a limited set of resources or noted that everyone should have access to health care services. One participant in CP said—

But, I think our system is broken; it's—there are people falling through the cracks. And we do have some social systems like Medicare and Medicaid, but there are people who fall through the cracks that can't afford to buy individual health care or through their employer or whatever. But they also don't qualify for Medicare and Medicaid and all those programs. And it—it's just fundamentally unfair. There are people who work really hard, and then they have to make choices like.... [Participant began to cry.] (CP group in Sacramento, CA).

Participants often addressed fairness in connection with who pays for medical costs. Many participants thought it was fair that wealthy people would have more provider or treatment choices. Notably, when choosing between health plans as the town council in the hospital quality case study, some participants, in their role as town councilor, would chose the local low-volume hospital for the town employees but stated that they personally would find the money to go to the high-volume hospital for the relevant procedures. Many participants extended this line of thinking as a rationale for the belief that insurers, employers, and government have the right, as payers, to restrict choice. In that context, it was important to participants that those established limits provide better care as well as lower costs.

Similarly, when facilitators prompted participants to consider whether people could pay an additional premium to preserve choice, many participants thought this was unfair. In this context, some participants thought it was unfair for employees to be burdened with extra cost so that *other people* would have more choice, which is in keeping with the general sense that fairness means individuals should pay to have more choice. However, others may simply have been expressing a more general frustration with premium costs (i.e., that it is unfair for premiums to be so high, yet seem to provide few benefits).

I have a really good doctor that I love and [who] works for a specific hospital, and I don't get her anymore. So to me, I feel like, Oh my God, that's not fair, but at the same time I understand there needs to be some boundaries. Otherwise, our insurance payments are high now, so I pay like \$600 a month. So that's ridiculous (BCD group in Chicago, IL).

One could also interpret participants' positive attitudes toward permeable boundaries as an element of fairness: allowing for exceptions is fair when there are exceptional circumstances. This aspect of fairness supports some freedom and personal choice.

Personal Responsibility

With individual freedom comes personal responsibility. Participants referenced patients' responsibility to make informed health care choices, ask questions of their doctors, and maintain a healthy lifestyle. Participants offered education, information access, and cost transparency as ways to enable and encourage personal responsibility. Often, participants also felt that people who make irresponsible choices should pay for those choices themselves:

My feeling has always been, I don't really care what people do with their lifestyle. If they choose to do something wild and crazy, they got to pay for it. You know if you choose to smoke, you're paying for it in your body, but I'm also paying for it.... Why am I helping subsidize your lifestyle, you know? You want to do that, go and pay for the whole thing (CD group in Chicago, IL).

Others acknowledged that people have the right to make “*stupid choices*,” even if those choices affect others in society.

Respect for Science

Although participants expressed some concerns about medical evidence and research, citing issues including profit motive and lack of participation of relevant populations, most participants appreciated the role of evidence in medical practice. In this way, respect for evidence supports individual freedom and choice, as doctors and patients could use evidence in making decisions. But they appeared to want and expect science to produce clear and compelling evidence before restricting choice or autonomy.

Pragmatism

Participants’ practical concerns about boundary implementation may reflect elements of pragmatism or practicality. Participants wanted to solve practice issues and ensure that boundaries were feasible and efficient before removing individual freedom and choice. For many, the logistical and practical challenges were their primary objection to boundaries.

Summary of the Public’s Response to the Deliberative Question

Here we summarize how participants’ response to the deliberative question changed over the course of the deliberations.

Initial Reactions to the Overarching Question: Where Did People Start?

When asked the overarching question, participants first focused on the concept of boundaries. Many of them initially interpreted boundaries as compulsory rules that restricted choices and allowed no exceptions, and most reacted negatively. Participants also questioned what was meant by “society.” They initially defined society as the government or a health insurance company, perceiving both types of organizations as enforcers of boundaries in health care.

Participants also focused on the nature and importance of the doctor-patient relationship in medical decisionmaking, often referring to the expertise of doctors and the desire to preserve the autonomy of individual patients and doctors to make decisions.

Participants rarely discussed the concept of evidence or questioned what the terms “evidence” or “effectiveness” meant when initially responding to the question.

Process of Deliberation: What Happened?

Over the course of the deliberative sessions, participants expressed and debated additional viewpoints. Participants discussed evidence presented to them during deliberation and also offered their experiences and knowledge. Deliberation often encouraged participants to shift how they used and applied the concepts in the deliberative question to decisionmaking by exploring different perspectives or viewpoints. Asking participants to take on a specific role as a societal decisionmaker (i.e., town council member) or to debate the merits of opposing positions was a principal driver to encourage exploration of different perspectives.

Discussions elicited other interpretations of “boundaries” and explored alternative approaches to getting patients and doctors to consider medical evidence, which was generally perceived as

desirable, in decisionmaking. Throughout, participants emphasized their desire for education and information about health and health care. Many participants presented education of patients, doctors, or both as the primary and preferred alternative to restrictive boundaries. They also articulated expectations that good doctors would know and use the most current evidence along with their own knowledge and experience, and would bring relevant evidence to their patients' attention. However, they also debated the use of rules, guidelines or standards, accountability mechanisms, and penalties or incentives.

Participants also deliberated on several concepts embedded in the overarching question, such as evidence of effectiveness, evidence of harm, and health care costs. In general, participants viewed evidence of effectiveness positively and expressed strong support for using evidence to make informed health care decisions. Participants weighed evidence of effectiveness with evidence of harms, often considering risks for individuals and communities. Participants discussed two primary types of harm: physical and economic. Physical harm could accrue both to individuals and to the community or society at large, while economic harm was discussed primarily in terms of the community. Further, participants raised cost issues spontaneously or addressed them in connection with discussions of boundaries and quality of care. Participants raised a variety of types of costs, both personal expenses and pooled resources such as insurance or tax funds, and also recounted difficulties they experience in determining what health care costs.

Final Reactions to the Overarching Question: Where Did People End?

Although many participants focused on how to preserve choice and enhance the doctor-patient relationship, the majority of participants came to the conclusion that some boundaries would be important or necessary as a means to address problems in the health care system or to bring about certain outcomes. When considering boundaries, relative assessments of risk of physical and economic harm often influenced attitudes about whether society should establish a boundary: the greater the risk, the more support for the boundary. Participants also considered harm that might result from ignoring or being unaware of evidence, or from having or not having boundaries on decisionmaking.

Most participants remained strongly committed to principles of freedom of choice and the autonomy of a trusted medical expert. At the same time, the debate over the acceptability of boundaries and consideration of evidence elicited other values, such as concerns for the greater good or fairness.

Discussion

As part of AHRQ's overall mission to improve the quality, safety, efficiency, and effectiveness of health care for all Americans, the Agency supports health services research that will promote evidence-based decisionmaking and improve the overall quality of health care. Below, we discuss these points:

- How the public input from the Community Forum can apply to these AHRQ focus areas
- The values elicited by deliberation
- How the process of deliberation affected public input

Promotion of Evidence-Based Decisionmaking

In considering personal medical decisions, participants' comments indicated that they placed high value on evidence of medical effectiveness and would want both certain and uncertain evidence to make informed personal decisions. Moreover, participants said that their access to easily understandable information about evidence is limited; they wanted more transparent information about quality and costs of health care services. These findings indicate strong support for research and efforts to translate research findings for consumer audiences and increase awareness of current resources.

Yet evidence was not always the primary factor in decisionmaking. Other factors such as personal preferences, their doctor's clinical judgment, and costs often outweighed the primacy of evidence. Participants often indicated they might be willing to consider treatments for which no evidence existed, especially as a last resort. Also, evidence of any kind might quickly be discounted if it was seen as "not applicable to me," indicating the need to include diverse populations in research studies and more clearly state when diverse populations are included in the evidence base.

Over the course of deliberation, participants discussed promoting evidence-based decisionmaking through boundary-setting mechanisms. Participants almost always preferred boundaries that preserve choice and the doctor-patient relationship. They actively sought alternatives to restrictive boundary setting across a variety of contexts—emphasizing doctor and patient education, proposing accountability measures, or proposing alternative approaches to problem solving than those used in the case studies. When participants considered evidence in the context of boundary setting, the type of evidence influenced the acceptability of different types of boundaries. By itself, evidence of medical effectiveness was not a compelling driver of acceptance of stricter boundaries. Evidence of physical and economic harm, however, did drive acceptance of stricter boundaries.

For example, participants who discussed only the hospital quality case study tended to prefer loose boundaries until reaching the point where economic harm to the community is introduced. At that point, the focus of discussion changed course, and comments indicated that participants supported a stricter boundary on choice. By contrast, participants who also discussed the URI case study, which emphasized physical harm resulting to society from ignoring evidence, appeared more accepting of rules and regulations than participants who did not discuss this case study. In addition, participants in the CP method, who discussed the heart disease case study, were the most likely to generalize that guidelines, as a boundary-setting approach, were ineffective and that regulations were necessary to achieve desired outcomes. In the CP method, facilitators assessed participants' attitudes toward societal boundaries on a daily basis. They detected the most shifts in perspective on the third day, after discussion of all individual case studies, when participants freely and openly discussed issues relevant to the deliberative topic in a technique called "open space." Although it is difficult to assess with certainty how these differences between cases and methods arose, it could be that differences in intensity—the time available to discuss multiple case studies, learn more information about current health care issues, explore alternative points of view, and interact with experts—allowed for a more nuanced discussion of the topic.

Finally, it is important to note that public perceptions of evidence appear to be broader than those of researchers, clinicians, or other experts in health care. Although participants discussed

evidence as the results of research studies in the context of the case studies, comments indicated that, at times, participants also equated evidence with experience—the doctor’s accumulated experience and clinical judgment, personal experiences, or common sense.

Improving Quality of Health Care

Discussions highlighted aspects of quality of care that were important to the public. For example, participants identified doctors as best equipped to understand and make decisions for individual patients based on evidence. But participants did not think that doctors were infallible and often called for increased oversight and accountability of doctors. Also, in the hospital quality case study particularly, participants debated the relative importance of clinical outcomes, such as survival, with patient-centeredness, such as being treated with respect. Although many prioritized evidence of clinical outcomes, others prioritized receiving personal and respectful care. Efforts to improve quality should include both aspects of care.

As AHRQ and others work to improve quality, it is also important to note that participants expressed concern that the motivation for establishing boundaries, or making any changes in access to health care, was cost containment rather than ensuring quality. In addition, participants’ initial negative reactions to boundaries may suggest an assumption or belief that health care is always beneficial and not harmful; boundaries are seen as preventing patients from getting care they may need rather than protecting people from harm. When participants took evidence of harm into account during deliberation, they offered different views and suggestions for the situation at hand, but also acknowledged that the problems facing the system are complex and require more discussions and deliberation. For AHRQ, understanding the context of the public’s concerns and participants’ initial reactions is important when communicating about evidence and quality improvement initiatives or assessing the impact of quality improvement measures. Also, a concerted effort by the health care system to engage the public in discussions about the integration of evidence, quality, and costs could produce actionable input for tailoring communication with patients. At the same time, the process of engaging the public also may help address fears about how evidence-based care is deployed and may help to provide the very type of health-related education and transparency that the public consistently desired.

Values Elicited by Deliberation

In discussions of evidence and boundaries, we identified values expressed by participants. The value of individual freedom was most often expressed, with values for the greater good and fairness arising with application of evidence of physical or economic harm. It is interesting to note how these values compare with an established framework of biomedical ethics, Beauchamp and Childress’ Four Principles:²⁰ respect for autonomy, beneficence, non-maleficence, and justice. Of utmost priority to deliberative participants were individual freedom and preservation of choice in medical decisionmaking, which align with the principle of autonomy. Also, participants’ emphasis on better education and information speaks to informed consent processes and procedures as a way to respect the autonomous choices of individuals. When faced with evidence of physical harm to individuals and the community, discussions moved to prioritizing the greater good, which highlights the principles of beneficence and non-maleficence. Finally, participants raised the value of fairness, aligning with the principle of justice, when assessing harms of allocating costs of health care services. This value was addressed to a lesser extent than

the other values, perhaps because cost and allocation of resources were not the primary focus of these deliberations.

Process of Deliberation Affecting Public Input

Deliberation allowed participants to explore their own views in more detail, to witness how information and context could shift the focus of discussions and elicit additional points of view, and to debate the applicability of all this to the question at hand. Deliberation did not, however, address all misperceptions of evidence or the health care system. For instance, some participants struggled with numeracy issues related to risk assessments, while others were not familiar with how research evidence is generated. But numerous participants commented, at the close of their sessions, that they had a deeper understanding of the issues and problems, and a better appreciation of a variety of factors relevant to health care.

Deliberation created a forum in which people expressed their views on a steadily shifting groundwork that required them to deal with a variety of tensions and factors inherent in the issue at stake. Some techniques used in the deliberative methods appeared to be effective at ensuring that participants grapple with these tensions and factors in answering the deliberative question. For example, asking participants to take on the role of societal decisionmaker (e.g., patient advisor to an insurance company, town council member) enabled them to shift positions simply by requiring them to consider the needs and concerns of others, or to bring different information or values to the fore. Also, asking participants to select among position statements engendered debate between participants and also debate about competing priorities within individuals. Many participants found it difficult to choose a single statement from the outset and provided new insights in the process of discussing how they would combine or rewrite the options; others chose different statements at different points in the deliberation, their selection influenced by the situations and roles to which they responded.

The ability of participants to hold apparently contradictory positions and values is a beautiful conundrum of our findings. This observation is borne out by the survey results discussed in subsequent chapters, in which groups can be seen shifting one direction on some factors or items and in the opposite direction on others, and where some methods seem to engender shifts in factors or items where, logically, they ought not to do so. The answer to these contradictions may lie in the process of deliberation. It does not merely ask participants to give their opinion, but frees them to shift their points of view, to play around with the factors involved, and to consider or even hold a variety of positions over the course of discussions.

Deliberation empowered participants to grapple with the entanglements and messy competing interests that exist in health care, in their relationships with their doctors, in their communities, and even within themselves. By doing so, it gave them an opportunity to explore more fully what was important, when it was important, who is involved, and the circumstances under which restrictions might be acceptable or even appropriate. The chosen question presented a difficult and complex issue—and deliberation allowed us to capture the complex response that emerged.

Chapter 5. Research Question 1: Comparing Effectiveness of Public Deliberation With Receiving Educational Materials

In this section, we report the quantitative results describing the effectiveness of deliberation. The intent of deliberation is to obtain *informed* public opinion; change in knowledge is thus a necessary, although not a sufficient, indicator of whether deliberation has occurred and a critical measure of the effectiveness of deliberation. A second indication that deliberation has occurred is a change in attitude about the topic of deliberation. A third component of the impact of deliberation is the quality of communication or discourse, reported in Chapter 6. The fourth component is the extent to which the deliberation elicited values through the discussion; this outcome was examined through the qualitative analysis presented in Chapter 4. For the purpose of this chapter, which presents solely the quantitative results, we present effectiveness of deliberation in terms of the impact it had on the knowledge and attitude measures employed here. We framed the general question about effectiveness as follows:

Is public deliberation more or less effective than educational materials alone at changing knowledge and attitudes about the deliberative topic, and is there a concomitant shift in attitudes?

We hypothesized that deliberation participants, regardless of the deliberation method, would demonstrate larger changes in knowledge and attitudes about the deliberative topics than control group members, as measured by the pre and post Knowledge and Attitudes Surveys. (See Chapter 3 for complete description of methods.) We also hypothesized that participants in each of the four deliberative methods individually would demonstrate larger changes in knowledge of and attitudes about the deliberative topics than control group members. The specific research questions (RQs) framed to address these hypotheses were:

RQ1.1. Do participants in all four deliberative methods combined demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

RQ1.2. Do participants in each of the four deliberative methods demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

To ensure that the differences observed between the experimental and control groups were attributable to whether or not they deliberated, rather than due to lack of access to information, we provided the control groups with case studies about three of the deliberative topics: *Upper Respiratory Infection in Children: Antibiotics Versus Symptom Management* (URI), *Obesity Management: Comparing Treatment and Prevention* (obesity management), and *Comparing Hospital Quality* (hospital quality). (See Chapter 2 for a full description of case studies.)

Methods

Sample and Power

An extensive discussion of the study recruitment and sampling strategy and overall participant characteristics is found in Chapter 3. With our sample size and allocation, the minimum detectable effect size (MDES) for the key comparisons ranges from 0.21 to 0.40, which is considered small to medium.¹⁷ In other words, this study is powered to find small differences between the pooled deliberation groups and the control group and medium differences between the individual methods and the control group. Medium differences are sufficient for this study because we are interested in important differences between methods, and while we were not able to detect small differences, small differences are less meaningful for informing the practice of deliberation.

Data Collection and Measurement

We measured changes in knowledge and attitudes using pre-deliberation and post-deliberation surveys. During the deliberative groups, participants were given specific examples in the form of case studies to illustrate the issues and promote deliberation. (See Chapter 2 for description of case studies.) Not all participants were exposed to each case study, and we elected a conservative approach to the analysis that does not assume spillover effects (i.e., changes in knowledge and attitude regarding a topic that was not discussed but might occur as a spillover following discussion of other related topics or concepts). Thus, topic-specific changes in knowledge and attitudes were assessed only for the topics each method discussed. A list of outcomes is presented in Exhibit 5-1 (Exhibits 5-1a through 5-1d).

The knowledge items measure knowledge of issues and concepts related to health care in the United States, the use of medical evidence, and comparative effectiveness research. Seven items were scored dichotomously as 1 = correct and 0 = incorrect. (See Exhibit 3-7 for a complete list of items.) Scores are reported as the proportion correct and range from 0 to 100 percent. The team designed the attitude items to measure attitudes surrounding the use of medical evidence in decisionmaking. We used factor analysis to define the structure of survey items and determine where it was appropriate to calculate summary scores for the domains of interest. Although we did not have a hypothesis about the directionality of shifts in attitudes, it was valuable to see if participants moved toward or away from the perspectives being measured. Thus we used relative, rather than absolute, values in all analyses. Because we assessed the impact on group means, our results capture net shifts in direction—i.e., the mean change in each group that occurred as a result of deliberation. (See Chapter 3 for a detailed description of the analytic approach.)

Exhibit 5-1a. Knowledge outcome measures

Knowledge	Outcomes Common to All Deliberative Methods
Knowledge of health care issues presented in educational materials	<ul style="list-style-type: none"> • The role of medical results in new medicine approval for use in the United States • The role of comparative effectiveness research in comparing medical treatments • Payment of health care costs in the United States • The role of comparative effectiveness research in developing quality standards for treatment • The effect of unnecessary care on health plan costs • The relationship between expense of medical treatment and effectiveness • Hospitals that perform many hip replacement surgeries each year get better results than hospitals that perform fewer hip surgeries

Exhibit 5-1b. Attitude outcome measures common to all deliberative methods

Attitude	Outcomes Common to All Deliberative Methods
Attitudes toward using medical evidence in health care decisionmaking	<ul style="list-style-type: none"> • Importance of knowing about medical evidence when making health care treatment decisions
	<ul style="list-style-type: none"> • Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking^a
	<ul style="list-style-type: none"> • Doctors and patients should consider evidence over preferences when making treatment decisions
Considerations of costs in health care decisionmaking	<ul style="list-style-type: none"> • Doctors and patients should consider cost evidence when making treatment decisions
	<ul style="list-style-type: none"> • People should consider the effect on group premiums when making treatment decisions^a
Attitude toward mechanisms to support use of high-volume hospitals	<ul style="list-style-type: none"> • Insurance companies should require children who need heart surgery to be treated at high-volume hospitals^a
	<ul style="list-style-type: none"> • Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children^a
	<ul style="list-style-type: none"> • Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results^a

^aIndicates single item used as outcome measure.

Exhibit 5-1c. Attitude outcome measures evaluated only for CD and CP

Attitude	Outcomes Evaluated Only for CD and CP
Attitude toward the use of medical evidence to limit antibiotic use	<ul style="list-style-type: none"> • People should not be able to get antibiotics if it may not help^a
	<ul style="list-style-type: none"> • Doctors should not be able to prescribe antibiotics if it may not help^a
	<ul style="list-style-type: none"> • The government should limit when doctors can prescribe antibiotics^a

^aIndicates single item used as outcome measure.

Note: CD = Community Deliberation; CP = Citizens' Panel.

Exhibit 5-1d. Attitude outcome measures evaluated only for CP

Attitude	Outcomes Evaluated Only for CP
Attitudes toward the role of society in prevention	• Role of community (parents, teachers, and schools) in limiting unhealthy food
	• Role of government policies in limiting unhealthy food
	• Role of government in encouraging (funding/incentives) preventive interventions to encourage healthy behaviors

Note: CP = Citizens' Panel.

Modeling

We evaluated whether mean pre-post changes of the outcome variables were equal across the different methods and control group while statistically controlling for the effects of the covariates using the analysis of covariance (ANCOVA) procedure in SAS version 9.2. We specified the following model:

$$KA_i = \beta_0 + \beta_1 pre\text{-}score + \beta_2 CP_i + \beta_3 CD_i + \beta_4 ODP_i + \beta_5 BCD_i + \delta X_i + \varepsilon_i$$

Where:

the unit of analysis is the individual participant;

KA_i is the change in knowledge or attitude for participant i ;

CP_i , CD_i , ODP_i , and BCD_i are dummy variables indicating the deliberative method to which participant i was assigned—Citizens' Panel, Community Deliberation, Online Deliberative Polling, or Brief Citizens' Deliberation (e.g., $CP_i = 1$ if participant i was assigned to the CP method and $CP_i = 0$ if not assigned to CP);

β_2 , β_3 , β_4 , β_5 are used to determine if the difference between the change in outcomes for each method is significant compared with the change in outcomes for the control;

δX_i is a vector of the control variables, including geographic location to control for regional variation in knowledge and attitudes, and the participant's health status, experience with the health care system, gender, age, marital status, education, employment status, and bilingual status to control for personal characteristics that might affect knowledge and attitudes; and ε represents the unexplained variance.

Weighting Procedure

The randomization process resulted in an approximately equal distribution of participants in each deliberative method; however, it produced a different distribution in the control group on (1) health insurance coverage, (2) income, and (3) racial/ ethnic groups. The difference occurred because we gave priority to assigning the priority population members to deliberation early in recruitment, meaning there was a higher proportion of the priority population groups in the deliberation arms than the control. To adjust for this imbalance, we calculated and applied analytic weights for analyses comparing the deliberative methods with the control group (RQ1). Participants were weighted to the U.S. Census population distribution of the relevant variables in the geographic location from which they were selected. See Appendix H for the detailed weighting procedure.

Clustering

Each deliberative session represents a cluster of participants who shared the deliberation experience; thus, we would expect that participants within a deliberative group may have had similar outcomes. This shared experience could have caused outcome variables to be more similar within these clusters than they would have been had each participant been able to deliberate independently, which is the underlying assumption of our ANCOVA models. It may be necessary to account for clustering by correcting the inaccurate standard errors caused by the intraclass correlation (ICC). A significant clustering effect would artificially reduce the variance and increase the chances that we would find a significant difference because of the clustering rather than the impact of deliberation. The typical solution to avoid this error is to adjust the variances for the degree to which outcomes reported by persons in the same deliberative group are correlated with each other, the ICC. We computed the ICCs of the knowledge and attitude outcomes, treating each deliberative group as a cluster. Because the ICCs were very close to zero (mostly below 0.02), the clustering effect is negligible, and thus we did not adjust the variances for ICC in RQ1 analyses.

Results

Exhibit 5-2 (Exhibits 5-2a through 5-2c) shows the comparison between changes in all the deliberative groups combined compared with the control group (RQ1.1). Exhibit 5-3 (Exhibits 5-3a through 5-3e) presents results from the individual methods compared with the control group (RQ1.2). We report an estimate of the difference between deliberation and control (change in mean outcome) and the associated significance level. We expand below when there was a significant effect of deliberation compared with the control group on the knowledge or attitude outcome variables when controlling for the covariates (pre-deliberation survey score, geographic location, health status, experience with health care system, gender, age, marital status, education, employment status, and bilingual status).

Effect of Deliberation on Knowledge and Attitudes

RQ1.1. Do participants in all four deliberative methods combined demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

Knowledge. The knowledge outcome captures the information gained based on questions that were linked to the background educational materials provided to all participants, including the control group. For the participants of deliberation, there were likely other knowledge gains, but our focus was to measure the information contained in the educational materials to conduct the most conservative test of the impact of deliberation on knowledge. All groups—the control group as well as the combined deliberative groups—increased in knowledge of issues and concepts covered in our survey, including health care in the United States, the use of medical evidence, and comparative effectiveness research (Exhibit 5-1). This change was measured from pre-test to post-test. Further, deliberation had a significant effect ($p = 0.01$) on the increase in knowledge for participants when compared with the control group and controlling for the covariates, demonstrating that a greater increase in knowledge occurred in the deliberative groups than in the control.

Attitudes toward using medical evidence in health care decisionmaking. Deliberation had a statistically significant effect on shift in participants' attitudes toward—

- *Importance of knowing about medical evidence when making health care treatment decisions* ($p < 0.01$)
- *Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking* ($p = 0.01$; single item)
- *Doctors and patients should consider cost evidence when making treatment decisions* ($p < 0.01$)

However, deliberation did not have a significant impact on the following outcomes related to the use of evidence:

- *Doctors and patients should consider evidence over preferences when making treatment decisions*
- *People should consider the effect on group premiums when making treatment decisions* (single item)
- *Insurance companies should require children who need heart surgery to be treated at high-volume hospitals* (single item)
- *Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children* (single item)
- *Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results* (single item).

Exhibit 5-2a. Effect of deliberation on knowledge: pooled deliberative groups versus reading materials only control

Outcomes	Method	Pre-Mean	Post-Mean	β	p-Value
Knowledge	Control	39.8%	47.9%	-	-
	All deliberation	38.3%	50.6%	-	-

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group.

Exhibit 5-2b. Effect of deliberation on attitudes toward using medical evidence in health care decisionmaking: pooled deliberative groups versus reading materials only control

Outcomes	Method	Pre-Mean	Post-Mean	β	p-Value
Importance of knowing about medical evidence when making health care treatment decisions	Control	4.4	4.4	-	-
	All deliberation	4.4	4.5	0.55	< 0.01 ^a
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking (single item)	Control	2.7	2.6	-	-
	All deliberation	2.6	2.8	0.64	0.01 ^a
Doctors and patients should consider evidence over preferences when making treatment decisions	Control	2.9	2.9	-	-
	All deliberation	2.9	2.9	0.33	0.08
Doctors and patients should consider cost evidence when making treatment decisions	Control	2.5	2.6	-	-
	All deliberation	2.5	2.7	0.83	< 0.01 ^a
People should consider the effect on group premiums when making treatment decisions (single item)	Control	2.8	2.9	-	-
	All deliberation	2.8	2.9	0.52	0.11

^aIndicates a significant difference between deliberative method and control at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group.

Exhibit 5-2c. Effect of deliberation on attitudes toward mechanisms to support use of high-volume hospitals: pooled deliberative groups versus reading materials only control

Outcomes	Method	Pre-Mean	Post-Mean	β	p-Value
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals (single item)	Control	3.1	3.2	-	-
	All deliberation	3.2	3.4	0.50	0.11
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children (single item)	Control	3.9	3.9	-	-
	All deliberation	3.9	3.9	0.10	0.71
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results (single item)	Control	4.1	4.2	-	-
	All deliberation	4.0	4.1	-0.11	0.60

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group.

Effect of Individual Methods on Knowledge and Attitudes

RQ1.2. Do participants in each of the four deliberative methods demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

Below we present results from the individual methods compared with the control group. Participants in different methods were exposed to varying numbers of case studies based on length of the method; thus results are presented only for the case studies presented in each group. The statistics reported estimate the difference between deliberation and control in change in mean outcome and the associated significance level.

Knowledge. Each group had notable increases in knowledge scores capturing issues and concepts related to health care in the United States, the use of medical evidence, and comparative effectiveness research. See Exhibit 5-3 (Exhibits 5-3a through 5-3c) for full results. The mean increases were, in order of magnitude of increase—

- Control, 8.1 percent
- ODP, 10.7 percent
- CD, 11.6 percent
- BCD, 12.5 percent
- CP, 17.2 percent

Deliberation had a significant impact on knowledge for participants in CP ($p = 0.01$) and BCD ($p = 0.02$), meaning that these groups saw statistically significant gains above what was observed in the control group.

Attitudes toward using medical evidence in health care decisionmaking. Compared with controls, deliberation had a significant impact on the shift in participants' attitudes toward:

- *Importance of knowing about medical evidence when making health care treatment decisions* for participants in three methods: CP ($p < 0.01$), CD ($p = 0.03$), and ODP ($p < 0.01$). This shift was significant even though all participants, even control, agreed very strongly with this view at the outset.
- *Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking* (single item) for two methods: CD ($p < 0.01$) and BCD ($p = 0.02$).
- *Doctors and patients should consider evidence over preferences when making treatment decisions* for CP ($p = 0.01$).

Attitude toward the use of evidence related to cost considerations in health care decisionmaking. Deliberation had a statistically significant impact on attitudes toward costs in decisionmaking for participants in the CP group only. Impacts were seen for both the measures:

- *Doctors and patients should consider cost evidence when making treatment decisions* ($p < 0.01$).
- *People should consider the effect on group premiums when making treatment decisions* (single item; $p < 0.01$).

Attitude toward mechanisms to support use of high-volume hospitals. In the hospital quality case study, discussed in all four methods, participants were asked to weigh the use of evidence in the context of impacts on the well-being of their community. Deliberation did not have a significant impact on participant weight of evidence related to the following measures: *insurance companies should require children who need heart surgery to be treated at high-volume hospitals* (single item), *hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children* (single item), and *doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results* (single item). Of note, there was strong agreement among all participants with this last measure, both before and after deliberation.

Attitude toward the use of medical evidence to limit antibiotic use. Three items were designed to measure if views shifted about the overuse of antibiotics for the two methods that discussed this case, CP and CD. These items reflect attitudes about three levels of limits on people's ability to obtain antibiotics, relative to a case describing harms of inappropriate antibiotic use: (1) general restrictions on people (patients), (2) general restrictions on doctors, and (3) government-level restrictions on doctors. There were significant shifts in attitudes in support of the role of government in setting limits on when doctors can prescribe antibiotics:

- *Government should limit when doctors can prescribe antibiotics* (single item; $p = 0.01$ for CP and $p < 0.01$ for CD)

but not for the other two measures, *people should not be able to get antibiotics if it may not help* (single item) and *doctors should not be able to prescribe antibiotics if it may not help* (single item).

Attitudes toward the role of society in prevention. The CP group was the only group that deliberated this topic. Deliberation did not have an impact on the measures *role of community (parents, teachers, schools) in limiting unhealthy food*, *role of government policies in limiting unhealthy food*, or *role of government in encouraging (funding, incentives) preventive interventions to encourage healthy behaviors*.

Exhibit 5-3a. Effect of deliberative methods on knowledge: each deliberative method versus control

Outcome	Method	Pre-Mean	Post-Mean	β	p-Value
Knowledge	Control	39.8%	47.9%	-	-
	CP	36.1%	53.3%	0.07	-
	CD	39.3%	50.9%	0.03	0.12
	ODP	36.6%	47.4%	0.00	0.79
	BCD	39.6%	52.2%	0.04	0.02 ^a

^aIndicates a significant difference between deliberative method and control at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 5-3b. Effect of deliberative methods on attitudes toward using medical evidence in health care decisionmaking: each deliberative method versus control

Outcome	Method	Pre-Mean	Post-Mean	β	p-Value
Importance of knowing about medical evidence when making health care treatment decisions	Control	4.4	4.4	-	-
	CP	4.4	4.6	0.27	< 0.01 ^a
	CD	4.4	4.5	0.11	0.03 ^a
	ODP	4.4	4.5	0.14	< 0.01 ^a
	BCD	4.4	4.4	0.04	0.51
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking (single item)	Control	2.7	2.6	-	-
	CP	2.6	2.8	0.17	0.12
	CD	2.6	2.8	0.22	< 0.01 ^a
	ODP	2.6	2.7	0.05	0.51
	BCD	2.6	2.8	0.19	0.01 ^a
Doctors and patients should consider evidence over preferences when making treatment decisions	Control	2.9	2.9	-	-
	CP	3.0	3.2	0.23	0.01 ^a
	CD	2.8	3.0	0.10	0.06
	ODP	2.9	2.9	0.00	0.98
	BCD	2.9	2.9	0.00	0.93
Doctors and patients should consider cost evidence when making treatment decisions	Control	2.5	2.6	-	-
	CP	2.5	3.2	0.59	< 0.01 ^a
	CD	2.5	2.7	0.08	0.26
	ODP	2.5	2.6	0.07	0.29
	BCD	2.6	2.7	0.09	0.24
People should consider the effect on group premiums when making treatment decisions (single item)	Control	2.8	2.9	-	-
	CP	2.7	3.3	0.48	< 0.01 ^a
	CD	2.8	3.0	0.08	0.42
	ODP	2.8	2.8	-0.09	0.36
	BCD	2.7	2.9	0.05	0.60

^aIndicates a significant difference between deliberative method and control at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 5-3c. Effect of deliberative methods on attitudes toward mechanisms to support use of high-volume hospitals: each deliberative method versus control

Outcome	Method	Pre-Mean	Post-Mean	β	p-Value
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals (single item)	Control	3.1	3.2	-	-
	CP	3.3	3.5	0.15	0.30
	CD	3.2	3.3	0.10	0.30
	ODP	3.3	3.5	0.14	0.15
	BCD	3.2	3.4	0.11	0.27
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children (single item)	Control	3.9	3.9	-	-
	CP	3.8	3.9	0.14	0.22
	CD	3.8	3.8	0.00	0.98
	ODP	4.0	4.0	0.09	0.25
	BCD	3.8	3.7	-0.14	0.11
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results (single item)	Control	4.1	4.2	-	-
	CP	4.0	4.2	0.00	0.99
	CD	4.0	4.1	-0.05	0.42
	ODP	4.1	4.2	0.03	0.68
	BCD	4.0	4.1	-0.08	0.24

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 5-3d. Effect of deliberative methods on attitudes toward the use of medical evidence to limit antibiotic use: each deliberative method versus control

Outcome	Method	Pre-Mean	Post-Mean	β	p-Value
People should not be able to get antibiotics if it may not help (single item)	Control	3.6	3.6	-	-
	CP	3.9	3.9	0.16	0.34
	CD	3.5	3.7	0.11	0.24
Doctors should not be able to prescribe antibiotics if it may not help (single item)	Control	2.4	2.6	-	-
	CP	2.6	2.8	0.19	0.17
	CD	2.5	2.7	0.17	0.06
Government should limit when doctors can prescribe antibiotics (single item)	Control	1.9	2.1	-	-
	CP	2.0	2.6	0.40	< 0.01 ^a
	CD	2.0	2.8	0.57	< 0.01 ^a

^aIndicates a significant difference between deliberative method and control at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. CD = Community Deliberation; CP = Citizens' Panel.

Exhibit 5-3e. Effect of deliberative methods on attitudes toward the role of society in prevention: each deliberative method versus control

Outcome	Method	Pre-Mean	Post-Mean	β	p-Value
Role of community (parents, teachers, schools) in limiting unhealthy food	Control	4.5	4.5	-	-
	CP	4.6	4.6	0.06	0.31
Role of government policies in limiting unhealthy food	Control	2.6	2.8	-	-
	CP	2.6	2.6	-0.16	0.07
Role of government in encouraging (funding, incentives) preventive interventions to encourage healthy behaviors	Control	3.3	3.4	-	-
	CP	3.5	3.4	-0.12	0.39

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. CP = Citizens' Panel.

Summary and Discussion

In this chapter, we address the overarching question of whether public deliberation is more or less effective than educational materials alone as measured by change in knowledge and attitudes about the specific topics of deliberation.

Impact of Deliberation Versus Control

RQ1.1. Do participants in all four deliberative methods combined demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

Pooling our deliberative methods to compare with control gave us our strongest measure of whether deliberative methods, taken as a group, have an impact on knowledge and attitudes compared with using educational materials alone.

Participating in deliberation increased participants' knowledge of evidence and comparative effectiveness research. Deliberation (for members of all groups combined) increased participants' knowledge about issues and concepts provided in the educational materials, including the use of medical evidence and comparative effectiveness research, as compared with the control group. Importantly, this increase in knowledge surpassed a statistically significant increase in knowledge that occurred within the control group following the control's review of educational materials. The survey questions designed to assess knowledge were based directly on what was presented in the materials; thus the increase in knowledge in the deliberative groups versus control represents a clear effect of deliberation on information gained and retained above the use of educational materials alone. There were likely other knowledge gains for participants in the deliberative groups based on the discussion that took place in their groups, but those were not captured in this survey.

Participating in deliberation shifted participants' attitudes regarding the role of evidence in decisionmaking but did not shift views regarding the relative importance of evidence and preferences. Deliberation (for members of all groups combined) shifted participants' attitudes related to the importance of medical evidence at a statistically significant level. Specifically, deliberation had an impact on *the importance of knowing about medical evidence when making health care treatment decisions* and attitudes toward *medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking*, but a shift did not occur in the measure *doctors and patients should consider evidence over preferences when making treatment decisions*. These items can be seen as hierarchical, in that one may be more willing to accept the importance of knowing about medical evidence, or even to be convinced that evidence is more important than a physician's knowledge of a patient, before one would accept that evidence is more important than preferences. Deliberation had an impact on the "lower bar" of perceived importance but stopped short of the strongest changes in attitude in this domain.

Impact of Individual Deliberative Methods Versus Control

RQ1.2. Do participants in each of the four deliberative methods demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

In these analyses, we evaluated the impact of each deliberative method compared with control. The level of power in our study permitted us to detect medium effect sizes. Thus, if individual methods were notably different (that is, meaningfully different) from controls, we detected this difference. If there were only small differences in impact, we would not detect them.

When comparing each deliberative method with the control group, all four deliberative methods showed significant change on at least one knowledge or attitude measure. On our knowledge measure, a statistically significant impact was seen for CP and BCD, but there was, in fact, an increase in mean knowledge scores in all groups. As discussed previously, the test of significance for this measure was very stringent, as the increase in knowledge for controls (based on the educational materials alone) was substantial: 8.1 percent.

There were changes in attitudes toward the use of medical evidence, an outcome central to our deliberative question, for all methods. As described above, this domain included outcomes that assessed the importance of *knowing* about medical evidence when making health care treatment decisions (shifts for CP, CD, and ODP); the importance of evidence *versus doctor's knowledge* about the patient when making health care treatment decisions (shifts for CD and BCD); and doctors and patients should consider medical evidence *over preferences* when making treatment decisions (shifts for CP). Again, the last measure may be the most difficult to shift.

In addition, the attitudes regarding use of medical evidence to limit antibiotic use, measured for two of our groups, CP and CD, show a similar impact of deliberation on attitude regarding use of evidence, namely a shift toward greater acceptance (or less disagreement) with a government role in constraining access to antibiotics. Findings from the qualitative data may help explain why participants shifted toward increased boundaries or accountability. Most discussions about government limits focused on adhering to guidelines and increased accountability of doctors. Participants generally interpreted guidelines as nonbinding recommendations that allowed clinicians to depart from these recommendations when appropriate. At the same time, participants often mentioned the need for doctors to verify test results of a bacterial infection or for an outside source to track the number of prescriptions being written for antibiotics.

Finally, compared with the control group, the CP method shifted participants' attitudes related to considering costs in making treatment decisions at a statistically significant level. Shifts showed increasing agreement with the factor *doctors and patients should consider cost evidence when making treatment decisions*. Participants in all methods received background materials about cost considerations in health care and discussed health care costs during deliberation. However, participants in the CP method also interacted with experts about health care costs and discussed costs in multiple case studies.

Shifts did not occur in three items related to the hospital quality case study that was used by all methods. There were no shifts at a significant level on attitudes related to the material in the hospital quality case study, which all the groups deliberated. This held true when comparing participants in all deliberative methods combined compared with the control group as well as when comparing participants in each method with the control group. The lack of significant findings for the measures of impact of deliberation on attitudes toward the hospital quality case study may be due to content and the complexity of this case study.

Initial discussions of this case often focused on whether evidence of clinical effectiveness should be the sole or primary basis for deciding on a hospital, or whether other aspects of health care, such as proximity to home and family or a better experience of care at a smaller hospital, were just as or more important. Although many participants argued that clinical outcome is the most important factor in quality health care, others prioritized receiving personal and respectful care. (The patient may be just a "*number*" at a high-volume hospital.)

When participants were asked to take on the role of town councilors in this case study, they weighed having access to the “better” hospital against the potential impact on the town of having a local hospital lose business and perhaps close because of reduced patient census. Evidence of clinical effectiveness did not support the importance of evidence, as it did in other cases. When taking a societal perspective, other factors besides clinical effectiveness needed to be considered.

These results present the effectiveness of each method in terms of quantitative changes in outcomes. However, to examine the overall impact of each method, we must consider all four indications of effectiveness: change in knowledge, change in attitudes about the topic of deliberation, the self-reported measure of quality of communication and discourse (reported in Chapter 6), and the values elicited through the discussion (reported in Chapter 4).

Chapter 7 presents the comparison of methods based on two main differences among deliberative methods, intensity and mode—differences sponsors may find particularly important to consider when choosing among deliberative approaches.

Chapter 6. Research Question 2: Participant Deliberation Quality and Experience

In this section, we report descriptive results of the quality of deliberative discourse and experience. In Chapter 7, we present additional deliberation quality and experience results when comparing the impact of method by intensity and mode.

The deliberative literature identifies the deliberation quality and experience as a primary measure of effectiveness of deliberation.¹ Quality of the deliberative experience may be defined as equal participation in the group, active participation from all participants, opportunity for discussion of the deliberative topic, and respect for other perspectives.

The considerable variation between deliberative methods in our study allows the opportunity to examine the effect of method on the deliberation quality and experience. The research question (RQ) addressing deliberation quality and experience was—

RQ.2. What was the overall quality of deliberative discourse and participant experience among the four methods?

We evaluated participants' experience during the deliberations on the basis of the quality of communication and discourse, the quality of the implementation process, and participants' ratings of the impact of deliberation.

Methods

Sample

The study recruitment and sampling strategy and overall participant characteristics are described in detail in Chapter 3.

Data Collection and Measurement

We created the Deliberation Quality and Experience Survey to measure three overarching domains: (1) discourse quality, (2) effect on participants, and (3) implementation process. Each domain includes a set of more narrowly defined outcomes. All outcome measures for these results are listed in Exhibit 6-1. A comprehensive list of all the outcomes and associated reliability coefficients is presented in Exhibit 3-7.

Exhibit 6-1. Deliberation quality and experience outcome measures

Topic	Outcome
Quality of communication and discourse	<ul style="list-style-type: none">• Respect for the opinions of others• Equal participation in the discussion• Reasoned justification of ideas (single item)• Speaking as much as wanted to in group (single item)
Implementation process	<ul style="list-style-type: none">• Implementation quality• Facilitator neutrality
Participant ratings of the impact of deliberation	<ul style="list-style-type: none">• Effect of deliberation on participants• Perceived value of the event

The deliberation quality and experience survey was administered at the conclusion of each deliberative group to measure the experience of the participants in the sessions. (Thus there are no pre-deliberation values.) Values were on a scale of 1 to 4, and we can interpret a mean of 2.5 as the midpoint.

Analysis

For this analysis, we calculated and report means for each measure by method. Comparison between certain methods is reported in Chapter 7.

Results

RQ.2. What was the overall quality of deliberative discourse and participant experience among the four methods?

Participants' mean responses were overall extremely high (Exhibit 6-2).

Quality of communication and discourse. The measures under this domain and ranges of average participant values for each method were as follows:

- *Respect for the opinions of others* ranged from 3.5 to 3.6.
- *Equal participation in the discussion* ranged from 2.2 to 2.5.
- *Reasoned justification of ideas* ranged from 3.2 to 3.4.
- *Speaking as much as wanted to in the group* ranged from 3.2 to 3.4.

Implementation process. Values for the quality of the implementation process were overall quite high, with the following mean ranges:

- *Implementation quality* values ranged from 3.2 to 3.5.
- *Facilitator neutrality* values ranged from 3.4 to 3.7.

Participant ratings of the impact of deliberation. The participants' perceived value of the impact of deliberation was similarly high on the two measures:

- *Effect of deliberation on participants* ranged from 3.0 to 3.4.

Perceived value of the event ranged from 3.5 to 3.7.

Exhibit 6-2. Deliberation quality and experience outcome measure mean scores for each method

Topic	Outcome	CP: Mean	CD: Mean	ODP: Mean	BCD: Mean
Quality of communication and discourse	Respect for the opinions of others	3.5	3.6	3.5	3.5
	Equal participation in the discussion	2.2	2.5	2.3	2.5
	Reasoned justification of ideas (single item)	3.4	3.4	3.2	3.3
	Speaking as much as wanted in group (single item)	3.3	3.4	3.4	3.2
Implementation process	Implementation quality	3.5	3.5	3.2	3.4
	Facilitator neutrality	3.4	3.4	3.7	3.4
Participant ratings of their experience in deliberation	Effect of deliberation on participants	3.4	3.3	3.0	3.2
	Perceived value of the event	3.7	3.7	3.5	3.6

Note: Possible values range from 1 to 4. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Discussion

Participants rated the quality of deliberation as high, both in terms of the quality of deliberative discourse and the implementation process. Participants across all methods rated the quality of communication and discourse highly. Participants agreed with the outcome measure that the participants in the groups showed respect for the opinions of others. Participants also rated agreement with the item that people gave reasons to support their ideas. Of note, participants' ratings for the factor *equal participation in the discussion* were relatively low compared with other measures of discourse quality; participants rated that some people in the group spoke more than others. Despite the fact that participants did not judge participation to be equal, they were still very satisfied with the experience.

Participants across all methods rated the implementation process highly. Values for the factor assessing the quality of the implementation process were overall high, including that the event was well organized, that the information presented was clear and easy to understand, and that the purpose of the event was clear. Also, values for the factor assessing facilitator neutrality were fairly high; participants in the Online Deliberative Polling (ODP) method rated *facilitator neutrality* notably higher than those in the other three methods. As described in Chapter 2, the neutrality of the group facilitator was a deliberate design characteristic of the methods; therefore, disparate outcomes were expected. Specifically, ODP facilitators were trained to intervene as little as possible during discussions, while still attempting to ensure that competing arguments in the reading materials were covered.

Participants reported that they placed a high value on taking part in deliberation and that the experience affected their opinions. High ratings of the factor *perceived value of the event* showed that participants found the activity to be valuable and indicated that they would like to participate in activities like this in the future. Stakeholder involvement research²¹ shows that participants are more likely to participate if they feel their input is being used in a meaningful way, and participants valuing the event may, in turn, make participants more likely to participate in future deliberative sessions. Ratings for the factor *effect of deliberation on participants* reflected participants' perceptions that the experience had an impact on their opinions on the deliberative topic.

Chapter 7. Research Question 3: Comparing Effectiveness of Specific Deliberative Methods

In this section, we discuss outcomes related to Research Question (RQ) 3: Are specific deliberative methods more effective than others? In Chapter 5 we examined the effectiveness of each method by comparing its impact with the impact of educational materials only on a group of outcome measures. Those results showed that each method had a significant impact on at least one outcome measure. As described in Chapter 2, the four methods for this experiment share certain core elements but also differ on key features of implementation, such as intensity and whether they take place in person or remotely. Research Question 3 focuses on these two main differences among the deliberative methods, because sponsors may find them particularly important to consider when choosing among deliberative approaches. The first, intensity, is the total amount of time spent in deliberation, which reflects the duration of deliberative sessions and number of sessions the method requires for full participation. The second, mode, is whether participation is in person or remote. The research team found these two elements to be conceptually the most important to assess based on previous research.¹

This analysis explores whether intensity and mode affect deliberation effectiveness by comparing the methods with each other on participant outcomes for knowledge of the deliberated topics, attitudes about the deliberated topics, and deliberation quality and experience. The analysis compares selected measures with each other; it does not compare all methods with each other or with the reading materials only control group.

RQ3.1. Does the intensity of the deliberative method affect knowledge, attitudes, and deliberation quality and experience?

The deliberative methods varied in length from 2 hours to 20 hours. With more time comes the opportunity to add content (e.g., more case studies), go into greater depth, engage with experts in the field, and deliberate in different ways (e.g., breakout groups, “open space”), as discussed in Chapter 2. Some methods also had more participants, exposing participants to more viewpoints. All these differences contribute to what we call *intensity*. This analysis compared the outcomes of participants in the method with the highest intensity, Citizens’ Panel (CP), to the method with the least intensity, Brief Citizens’ Deliberation (BCD).

We hypothesized that participants in the most intense method would experience larger changes in knowledge and attitudes about the deliberated topics and rate deliberation quality and experience more highly than participants in the least intense method because of the additional opportunity for engagement and time available for thoughtful debate.

RQ3.2. Does the mode of deliberation affect knowledge, attitudes, and deliberation quality and experience?

Online deliberation is an attractive alternative to in-person methods because of the cost associated with convening groups of people, the convenience for participants (including the potential to involve people who might not otherwise be able to participate), and the ability to bring together people living in different locations. However, some have raised concerns that online methods may be less conducive to deliberative discussion, and therefore less effective in eliciting values than in-person approaches.

This research question examines whether the effectiveness of online deliberation is similar to that of an in-person method. We hypothesized that participants in the exclusively online method, Online Deliberative Polling (ODP), would demonstrate similar changes in knowledge and attitudes about the deliberated topics and rate deliberation quality and experience similarly compared with participants in the primarily in-person method, Community Deliberation (CD), that had similar intensity.

Ideally, we would have addressed this question by comparing two methods that were identical except for the mode of deliberation. Our demonstration, which relied on existing deliberative methods as the basis for those included in the study, did not have two such methods. We therefore addressed this question by comparing the two methods that differed in mode but were closest in intensity, in an effort to hold constant this potentially important variable. (Intensity, as noted above, captures duration and other aspects of deliberative content.) It is important to note, however, that our online method, ODP, differed from the in-person comparator (CD) in facilitation style as well as mode. In addition, CD was not strictly an in-person method, as it incorporated an optional asynchronous online component between the two in-person sessions.

Methods

Sample

The statistical analysis for RQ3 included participants who attended the deliberative session(s) and completed both the pre and post Knowledge and Attitudes Surveys (for analyses related to knowledge and attitude outcomes) or completed both the pre Knowledge and Attitudes Survey and the Deliberation Quality and Experience Survey (for analyses related to experience outcomes). To ensure that the analysis sample was randomized and participants were comparable between the two pertinent methods in a comparison, we restricted the analysis sample for RQ3.1 and RQ3.2 to participants who were willing and available to participate in both pertinent methods in comparison and who were randomly assigned to either of the two methods. Thus, the participants used in a comparison were comparable across methods, and weights were not necessary for RQ3 analyses.

Exhibit 7-1 lists the number of participants included in each analysis. The detailed recruitment and sampling strategy and overall participant characteristics are described in Chapter 3. With our sample size and allocation, we had a medium minimum detectable effect size (MDES) for the comparisons¹⁷ (i.e., these analyses were powered to detect medium differences between methods but not to detect small differences).

Exhibit 7-1. Sample sizes for RQ3 analyses

Analyses	Effect of Deliberation Intensity: BCD	Effect of Deliberation Intensity: CP	Effect of Deliberation Intensity: Total	Effect of Deliberation Mode: CD	Effect of Deliberation Mode: ODP	Effect of Deliberation Mode: Total
Knowledge and attitudes outcome analyses	n = 100	n = 87	n = 187	n = 165	n = 163	n = 328
Deliberation quality and experience outcome analyses	n = 102	n = 86	n = 188	n = 164	n = 154	n = 318

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Data Collection and Measurement

The Knowledge and Attitudes Survey contained seven knowledge items on the deliberated topics. These were scored dichotomously as 1 = correct and 0 = incorrect and reported as a percent correct. The attitude items in this survey measure attitudes surrounding the use of medical evidence in decisionmaking. We used factor analysis to define the structure of survey items and determine where it was appropriate to calculate summary scores for the domains of interest. A total of three factors and five individual items were included as attitude outcomes in these analyses (Exhibit 7-2). We report relative values for the shifts in attitudes in order to show the directionality of the shifts, although, we did not have a hypothesis about the directionality of the impacts. (See Chapter 3 for a detailed description of the analytic approach.) The Deliberation Quality and Experience Survey measured three overarching domains: (1) discourse quality, (2) implementation process, and (3) effect on participants. A comprehensive list of all the outcomes and associated reliability coefficients is presented in Exhibit 3-7. All outcome measures for the intensity and mode comparisons are listed in Exhibit 7-2.

Exhibit 7-2. Knowledge, attitude, and deliberation quality and experience outcome measures

Topic	Outcome
Knowledge of health care issues presented in educational materials	<ul style="list-style-type: none"> • The role of medical results in new medicine approval for use in the United States • The role of comparative effectiveness research in comparing medical treatments • Payment of health care costs in the United States • The role of comparative effectiveness research in developing quality standards for treatment • The effect of unnecessary care on health plan costs • The effect of expense of medical treatments on efficacy • Hospitals that perform many hip replacement surgeries each year get better results than hospitals that perform fewer hip surgeries
Attitudes	
Attitudes toward using medical evidence in health care decisionmaking	<ul style="list-style-type: none"> • Importance of knowing about medical evidence when making health care treatment decisions • Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking (single item) • Doctors and patients should consider evidence over preferences when making treatment decisions
Considerations of costs in health care decisionmaking	<ul style="list-style-type: none"> • Doctors and patients should consider cost evidence when making treatment decisions • People should consider the effect on group premiums when making treatment decisions (single item)
Attitude toward mechanisms to support use of high-volume hospitals	<ul style="list-style-type: none"> • Insurance companies should require children who need heart surgery to be treated at high-volume hospitals (single item) • Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children (single item) • Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results (single item)
Deliberation Quality and Experience	
Quality of communication and discourse	<ul style="list-style-type: none"> • Respect for the opinions of others • Equal participation in the discussion • Reasoned justification of ideas (single item) • Speaking as much as wanted to in group (single item)
Implementation process	<ul style="list-style-type: none"> • Implementation quality • Facilitator neutrality
Participant ratings of the impact of deliberation	<ul style="list-style-type: none"> • Effect of deliberation on participants • Perceived value of the event

Modeling

For the knowledge and attitudes analysis, we evaluated whether mean pre-post changes in the outcome variables were equal between the two pertinent methods in a comparison, while statistically controlling for the effects of the covariates using the analysis of covariance (ANCOVA) procedure in SAS version 9.2.

$$KA_i = \beta_0 + \beta_1 \text{pre-score} + \beta_2 D_i + \delta X_i + \epsilon_i$$

Where:

KA_i is the change in knowledge or attitude for participant i ;

D_i is an indicator that = 0 if participant i was from CP (ODP in mode analysis) and 1 if participant was from BCD (CD in mode analysis);

β_2 is used to determine if the change in outcomes for CP is significantly different from the change in outcomes for BCD in intensity analysis or if the change in outcomes for ODP is significantly different from CD in mode analysis;

Pre-score is the pre-deliberation score of knowledge or attitude for participant i ;

X_i is a vector of the control variables, which included pre-deliberation survey score, geographic location, health status, experience with health care system, gender, age, marital status, education, employment status, health insurance coverage, income, race, ethnicity, and bilingual status; and

ϵ_{ij} represents the unexplained variance.

For the deliberation quality and experience outcomes, no pre-test was administered. Thus we evaluated the differences between the comparison methods for each of the eight deliberation quality and experience outcomes.

$$E_{ij} = \beta_0 + \beta_1 D_{ij} + \delta X_{ij} + \epsilon_{ij}$$

Where:

E_{ij} is the outcome measure of the deliberation quality and experience outcome for participant i in deliberative group j ;

D_{ij} is an indicator that = 0 if participant i was from CP (ODP in mode analysis) and 1 if participant was from BCD (CD in mode analysis);

X_{ij} is a vector of the control variables, including geographic location, health status, experience with health care system, gender, age, marital status, education, employment status, health insurance coverage, income, race, ethnicity, and bilingual status; and

ϵ_{ij} represents the unexplained variance.

Clustering

Treating each deliberative group as a cluster, the intraclass correlations (ICCs) of the knowledge and attitude outcomes were calculated. Because the ICCs were very close to zero (mostly below 0.02), the clustering effect is negligible and thus not considered in RQ3 analyses for knowledge and attitudes outcomes. However, calculation of the ICCs within deliberative groups revealed that within-group responses on the Deliberation Quality and Experience survey were correlated. This was not unexpected, given that participants shared their group experience. For these analyses we used generalized estimating equations (GEE), a process that extends the general linear regression model, takes clustering into account, and adjusts the resulting standard errors appropriately. Responses in a sample of participants where there is a clustering effect cannot be considered independent and thus violate the assumption of independence for regression models. Using the traditional linear regression models under such cases will lead to inaccurately smaller estimates of standard errors. Controlling for clustering adjusts for this and is thus a more conservative test.

Results

The results reported below indicate when one method of deliberation had a significantly different effect on the knowledge or attitude outcome variable compared with another method when accounting for covariates. Covariates included in the analysis were pre-deliberation survey score, geographic location, health status, experience with health care system, gender, age, marital status, education, employment status, health insurance coverage, income, race, ethnicity, and bilingual status. The β statistics estimated the difference between the two methods' change in means for knowledge and attitude outcomes or estimated the difference between the two methods' means for deliberation quality and experience outcomes. We also report below the significance levels associated with the β statistics and the group means for each outcome.

The knowledge outcomes are presented as percent correct. The attitude outcomes were scored on a five-point scale; Factors 2 and 3 and Items 12, 34, and 35 were recoded so that higher mean scores indicate a shift in attitude toward affirmative answers to the question. The deliberation quality and experience outcomes were scored on a four-point scale, with higher scores indicating better self-rating of deliberation quality and experience. On all outcomes, a positive coefficient indicates the score was higher in BCD than CP in the intensity analysis, or higher in CD than ODP in the mode analysis. Similarly, a negative coefficient means the average score on the outcome was lower for BCD than CP in the intensity analysis, or lower for CD than ODP in the mode analysis.

Intensity

RQ3.1. Does the intensity of the deliberative method affect knowledge, attitudes, and deliberation quality and experience?

Knowledge. We hypothesized that participants in higher intensity methods would demonstrate larger changes in knowledge of the deliberative topics than participants in lower intensity methods. The change in knowledge for the method with the lower intensity (BCD) was 6.6 percent, compared with 11.1 percent for the higher intensity method. However, this difference was not statistically significant. Of note, both methods had statistically significantly greater changes in knowledge than the control group. (See Chapter 5.)

Attitudes. Similarly, we hypothesized that participants in higher intensity methods would demonstrate greater changes in attitudes about the deliberative topics than participants in lower intensity methods. There were statistically significantly greater effects of intensity of deliberation in shifting responses to the following outcomes:

- *Importance of knowing about medical evidence when making health care treatment decisions* ($\beta = -0.25$, $p < 0.01$)
- *Doctors and patients should consider cost evidence when making decisions* ($\beta = -0.31$, $p = 0.03$)
- *Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children* (single item; $\beta = -0.31$, $p = 0.04$)

For these outcomes, both CP and BCD shifted in the direction of placing higher value on medical evidence, but the magnitude of movement was larger in the higher intensity CP. Shifts on the last item should be interpreted tentatively, because neither method had significantly greater shifts in

attitude than the reading materials only control group. (See Chapter 5.) While some descriptive differences can be seen in the means in Exhibit 7-3, there were no additional statistically significant differences between the methods on shifts in responses to the remaining measures:

- *Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking* (single item)
- *Doctors and patients should consider evidence over preferences when making treatment decisions*
- *People should consider the effect on group premiums when making treatment decisions* (single item)
- *Insurance companies should require children who need heart surgery to be treated at high-volume hospitals* (single item)
- *Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results* (single item)

Deliberation quality and experience. We hypothesized that participants in the most intense method, CP, would report more positive deliberation quality and experience than those in the least intense method. The deliberation quality and experience outcomes (means) for the CP method were higher than those for BCD with statistical significance for—

- *Implementation quality* ($\beta = -0.11$, $p < 0.01$)
- *Effect of deliberation on participants* ($\beta = -0.18$, $p < 0.01$)

However, deliberation quality and experience outcomes were higher for the lower intensity BCD method on two measures—

- *Equal participation in the discussion* ($\beta = 0.35$, $p < 0.01$)
- *Facilitator neutrality* ($\beta = 0.14$, $p = 0.02$)

The remaining outcomes—*respect for the opinions of others*, *reasoned justification of ideas* (single item), *speaking as much as wanted to in group* (single item), and *perceived value of the event*)—did not show a statistically significant difference.

Exhibit 7-3. Effect of deliberation intensity on knowledge, attitudes, and deliberation quality and experience (CP versus BCD; CP as reference)

Topic	CP: Pre-Mean	CP: Post-Mean	BCD: Pre-Mean	BCD: Post-Mean	β	p-Value
Knowledge of health care issues presented in educational materials	37.2%	52.4%	41.6%	49.9%	-0.04	0.15
Attitudes Toward Using Medical Evidence in Health Care Decisionmaking						
Importance of knowing about medical evidence when making health care treatment decisions	4.4	4.6	4.5	4.4	-0.25	< 0.01 ^a
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking (single item)	2.5	2.7	2.7	2.8	-0.11	0.44
Doctors and patients should consider evidence over preferences when making treatment decisions	3.1	3.2	3.0	3.1	-0.09	0.42
Considerations of Costs in Health Care Decisionmaking						
Doctors and patients should consider cost evidence when making treatment decisions	2.4	2.9	2.3	2.6	-0.31	0.03 ^a
People should consider the effect on group premiums when making treatment decisions (single item)	2.7	3.1	2.8	2.9	-0.30	0.07
Attitude Toward Mechanisms To Support Use of High-Volume Hospitals						
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals (single item)	3.6	3.5	3.2	3.3	-0.12	0.47
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children (single item)	3.9	4.0	3.9	3.8	-0.31	0.04 ^a
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results (single item)	4.0	4.2	4.1	4.2	-0.06	0.62
Quality of Communication and Discourse						
Respect for the opinions of others	-	3.5	-	3.5	0.02	0.70
Equal participation in the discussion	-	2.2	-	2.5	0.35	< 0.01 ^a
Reasoned justification of ideas (single item)	-	3.4	-	3.3	-0.04	0.42
Speaking as much as wanted to in group (single item)	-	3.3	-	3.3	0.01	0.94
Implementation Process						
Implementation quality	-	3.5	-	3.4	-0.11	< 0.01 ^a
Facilitator neutrality	-	3.3	-	3.5	0.14	0.02 ^a
Participant Ratings of the Impact of Deliberation						
Effect of deliberation on participants	-	3.4	-	3.2	-0.18	< 0.01 ^a
Perceived value of the event	-	3.7	-	3.7	-0.03	0.17

^aIndicates a significant difference between CP and BCD; CP as reference at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. BCD = Brief Citizens' Deliberation; CP = Citizens' Panel.

Mode

RQ3.2. Does the mode of deliberation affect knowledge, attitudes, and deliberation quality and experience?

Knowledge and attitudes. We hypothesized that participants in the exclusively remote method (ODP) would demonstrate changes in knowledge and attitudes about the deliberated topics similar to those of participants in the primarily in-person method (CD), which involved a similar amount of time spent in deliberation. Changes in knowledge and attitudes were not significantly different between the two methods for any outcomes. Pre and post means are reported in Exhibit 7-4.

Deliberation quality and experience. We similarly hypothesized no differences in deliberation quality and experience for the online versus the in-person methods. However, we did observe significant differences by mode. For six of the eight deliberation quality and experience outcomes, participants reported higher deliberation quality and experience for the in-person CD method than the ODP method:

- *Respect for opinions of others* ($\beta = 0.16$, $p < 0.01$)
- *Reasoned justification of ideas* (single item; $\beta = 0.26$, $p < 0.01$)
- *Speaking as much as wanted to in group* (single item; $\beta = 0.26$, $p < 0.01$)
- *Implementation quality* ($\beta = 0.30$, $p < 0.01$)
- *Effect of deliberation on participants* ($\beta = 0.25$, $p < 0.01$)
- *Perceived value of event* ($\beta = 0.20$, $p < 0.01$)

There was no difference between the methods in *equal participation in the discussion*. However, participants in the ODP method reported significantly higher outcomes for *facilitator neutrality* ($\beta = -0.20$, $p < 0.01$).

Exhibit 7-4. Effect of deliberation mode on knowledge, attitudes, and deliberation quality and experience (CD versus ODP; ODP as reference)

Topic	ODP: Pre-Mean	ODP: Post-Mean	CD: Pre-Mean	CD: Post-Mean	β	p-Value
Knowledge of health care issues presented in educational materials	34.7%	46.8%	39.4%	50.0%	0.01	0.54
Attitudes Toward Using Medical Evidence in Health Care Decisionmaking						
Importance of knowing about medical evidence when making health care treatment decisions	4.4	4.5	4.4	4.5	-0.04	0.48
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking (single item)	2.7	2.8	2.6	2.8	0.03	0.76
Doctors and patients should consider evidence over preferences when making treatment decisions	2.9	2.9	2.8	3.0	0.11	0.10
Considerations of Costs in Health Care Decisionmaking						
Doctors and patients should consider cost evidence when making treatment decisions	2.4	2.6	2.5	2.6	-0.05	0.56
People should consider the effect on group premiums when making treatment decisions (single item)	2.8	2.8	2.9	2.9	0.05	0.69
Attitude Toward Mechanisms To Support Use of High-Volume Hospitals						
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals (single item)	3.3	3.5	3.3	3.5	-0.07	0.57
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children (single item)	4.0	4.0	3.9	3.9	-0.06	0.49
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results (single item)	4.0	4.3	4.0	4.1	-0.15	0.07

Exhibit 7-4. Effect of deliberation mode on knowledge, attitudes, and deliberation quality and experience (CD versus ODP, ODP as reference) (continued)

Topic	ODP: Pre-Mean	ODP: Post-Mean	CD: Pre-Mean	CD: Post-Mean	β	p-Value
Quality of Communication and Discourse						
Respect for the opinions of others	-	3.5	-	3.6	0.16	0.01 ^a
Equal participation in the discussion	-	2.4	-	2.5	0.06	0.36
Reasoned justification of ideas (single item)	-	3.1	-	3.4	0.26	< 0.01 ^a
Speaking as much as wanted to in group (single item)	-	3.7	-	3.4	0.26	< 0.01 ^a
Implementation Process						
Implementation quality	-	3.2	-	3.5	0.3	< 0.01 ^a
Facilitator neutrality	-	3.6	-	3.4	-0.2	< 0.01 ^a
Participant Reports of the Impact of Deliberation						
Effect of deliberation on participants	-	3.0	-	3.3	0.25	< 0.01 ^a
Perceived value of the event	-	3.5	-	3.7	0.2	< 0.01 ^a

^aIndicates a significant difference between CD and ODP; ODP as reference at $p < 0.05$.

Note: Pre-mean indicates the mean score prior to deliberation; post-mean indicates mean score post deliberation; the estimate is the difference between the pre-post change for the deliberative groups and the pre-post change for the control group. CD = Community Deliberation; ODP = Online Deliberative Polling.

Discussion

Intensity

Intensity—as measured by contrasting the CP and BCD methods—did not increase knowledge but shifted attitudes at a statistically significant level. The higher intensity method (CP) did not increase participants' knowledge of evidence and comparative effectiveness research at a statistically significant level when compared with the lower intensity method (BCD). This result was counter to our hypothesis. However, it is important to note that across-the-board gains in knowledge were observed in our deliberative groups relative to the control groups (who received reading materials only). Knowledge scores for the highest intensity method (CP) were, in fact, the highest we observed (Chapter 5), but our findings suggest that the difference in method intensity does not achieve an increase in knowledge that is statistically significantly greater than what can be accomplished in carefully planned, less-intensive methods.

Participants in the higher intensity method (CP) were more likely to shift toward the consideration of evidence and costs in decisionmaking than participants in the lower intensity method (BCD). Intensity shifted participants' attitudes related to the importance of medical evidence on one factor, *the importance of knowing about medical evidence when making health care treatment decisions*, a measure central to our deliberative question. However, intensity did not significantly affect the factor *medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking* or the item *doctors and patients should consider evidence over preferences when making treatment decisions*. Participants in the CP method also had a greater change in attitude on the outcome *doctors and patients should*

consider cost evidence when making treatment decisions. This may have occurred because written cost considerations were included in three CP case studies and explicitly discussed within the groups—an extended conversation that was manageable in the longer high-intensity method. In most of the BCD groups, the facilitator introduced questions (probes) about cost informally as a topic of conversation, and discussion was necessarily constrained by the shorter time available.

Participants in the higher intensity method (CP) were more likely to consider evidence in the role of hospital quality. The CP group saw a greater shift in the item *hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children* than the lower intensity BCD group. This shift was consistent with CP participants' shift on *importance of knowing about medical evidence when making health care treatment decisions* (above). Of note, BCD participants actually shifted in the opposite direction—toward *disagreement* regarding the use of measures of success rates as a requirement for hospitals to perform heart surgery—following deliberation. Also, although we found a difference between CP and BCD on this measure, neither the shifts among participants in the CP nor the shifts in the BCD groups were statistically different from those in the control groups.

Intensity had an effect at a statistically significant level on participants' self-reports of the perceived value of the deliberative experience and the quality of communication and the implementation process. Both groups reported generally positive experience on the deliberation quality and experience measures (Chapter 6). Participants in the higher intensity method (CP) reported a greater impact from the experience than participants in the lower intensity method (BCD) reported, and this difference was at a statistically significant level. Participants in both methods rated the quality of deliberative discourse and implementation highly, but differed at a statistically significant level for three outcomes:

- Participants in the lower intensity method (BCD) reported more agreement with the two factors measuring equal participation in the discussion and facilitator neutrality than participants in the higher intensity method (CP).
- Participants in the higher intensity method (CP) reported higher ratings of the quality of the implementation process than participants in the lower intensity method (BCD).

Mode

Mode—as measured by contrasting the CD and ODP methods—did not change knowledge or attitude at a statistically significant level. Participants in CD, the in-person method, reported greater deliberation quality and experience than those participating in ODP, the remote method. Our comparison of an in-person (CD) versus online (ODP) method of similar intensity did not show a statistically significant effect on any of the knowledge or attitude outcomes. However, our comparison showed dramatic differences between these methods in deliberation quality and experience. Participants in CD reported significantly higher scores (at or below a $p = 0.01$ level) than ODP participants for six out of the eight measures of deliberation quality and experience.

One item, *facilitator neutrality*, was judged higher in our online method (ODP) than in the in-person method (CD). ODP may have been more successful in achieving facilitator neutrality in part because of the particular nature of our online method, in which facilitation was less active. This aspect of the method arguably had a greater impact on facilitator neutrality than the mode characteristic. However, this finding demonstrates the potential of the online method to achieve

greater success on this measure, and is an important signal for those implementing deliberative processes.

Our conclusion regarding the impact of mode is that remote methods have the potential to be similarly effective in affecting knowledge and attitudes as in-person methods of similar intensity, at least across the middle-intensity methods (5–6 hours of deliberation) that we examined. They are thus a credible alternative among deliberative methods and may be particularly useful in situations where gathering people in an in-person venue is difficult or impractical. In the methods we compared, a characteristic of ODP contributed to a less positive experience for participants. We are unable to determine if this characteristic was that ODP was online or that it used passive facilitation. Nonetheless, the strength of these results warrants caution regarding the deliberation quality and experience of online deliberation.

Chapter 8. Research Question 4: Examining Effectiveness of Public Deliberation by Participants' Personal Characteristics

In this section, we examine measured changes in knowledge and attitudes, and the perceived impact of deliberation, to answer Research Question (RQ) 4: Does the effectiveness of public deliberation vary by participants' personal characteristics?

In previous sections we describe the premises of deliberation and its goals for bringing informed public input to policy and programmatic decisions. One critical requirement of deliberative sessions is that they include representation from all parts of the community.¹² Participation by community members with a range of backgrounds allows all participants in deliberation to gain a better understanding of diverse opinions and perspectives on the programs or policies under study.

An important question regarding inclusion of different demographic groups in deliberation, however, is whether deliberation is similarly effective with all demographic groups. Does the effectiveness of deliberation—as measured by changes in knowledge and attitudes, as well as perceived impact—vary systematically by personal characteristics? We posit that consistent and systematic differences between demographic groups on these outcomes would suggest limited value of this method because they might indicate that some segments of the population are not engaged in the process.

This question has particular import in light of current efforts to engage audiences fully representative of their communities in discussions about health care and to encourage broad involvement in the development of health care research, policies, and programs. Examples of these efforts can be found in the areas of culturally and linguistically appropriate services, emergency preparedness, priority setting, and provision of culturally competent care.²²⁻²⁴ If deliberation is effective in increasing knowledge and encouraging reflection among all participants, it holds promise as a means for engaging different demographic groups and for assuring the representation of all demographic groups in the input that is generated.

The effectiveness of deliberation relies on maintaining a group dynamic conducive to all participants being able to express their views and deliberate on the topics presented. Some have questioned explicitly whether deliberation can achieve equal opportunity for participation: the suggestion is that people with more privileged status (e.g., education, income) may dominate deliberative discussions, leaving members of historically underrepresented groups less opportunity or less ability to participate.²⁵ (See also Chapter 9.) A variant on this argument is that more concordant groups, groups with participants of similar personal characteristics, offer a more comfortable environment to share values with others who have had similar experiences, and thus less concordant groups would inhibit free exchange of views. This contrasts with the intent of promoting diversity within deliberative groups and building on this diversity to ensure that participants are exposed to and discuss a wide range of experience, values, and opinions.

To explore the effectiveness of deliberation with respect to a variety of audiences, we evaluated (1) whether the effect of public deliberation on knowledge and attitudes varied by participants' personal characteristics, (2) how members of different demographic groups perceived the impact

and value of the experience, and (3) how group composition (concordance) affected these outcomes.

We examined the following questions:

RQ4.1 (Personal Characteristics). Do the deliberative outcomes (knowledge, attitudes, perceived impact) of African-American participants, Hispanic participants, participants age 65 and older, and participants without post-secondary education differ from the outcomes of others?

RQ4.2 (Concordance). Does group concordance, or the proportion of participants in the deliberative session who belonged to the same historically underrepresented group, affect deliberation outcomes for African-American participants, Hispanic participants, and participants without post-secondary education?

Methods

Sample

The recruitment, sampling, and randomization strategy for the study overall is described in detail in Chapter 3. Our strategy involved random allocation of willing participants to the deliberative methods. Our goal was to reflect the population's distribution in each of our four locations for gender, African-American race, Hispanic ethnicity, and age. To assess how well we achieved this goal, we compared study participants' demographic characteristics with the U.S. Census population characteristics for the four locations using a one-sample test of proportions. This test allows us to assess whether the study's sample mean significantly differs from the hypothesized value from the U.S. Census. There were 907 people recruited to deliberative discussion methods who participated in at least one discussion session. This count excludes people who were late, turned away, dropped out before the post-intervention survey, or were assigned to the control group.

We compare below and in Exhibit 8-1 (Exhibits 8-1a through 8-1d) the sample and the population for each demographic characteristic. The last column in Exhibit 8-1, labeled Total, reports on our goal of reflecting the population's distribution in locations. We also show the composition of the deliberative methods in each location.

Gender

The study sample was similar on gender in Durham, NC; Silver Spring, MD; and Sacramento, CA, but the study sample in Chicago, IL, had significantly more females compared with the population in Chicago.

Age

The study sample was similar on age in all four locations.

Race

The study sample in Durham, NC, and Silver Spring, MD, was similar on race to the populations in these locations. The study samples in Chicago, IL, and Sacramento, CA, had significantly more African-American participants than the populations.

Ethnicity

The study sample in Silver Spring, MD, and Sacramento, CA, was similar on ethnicity to the populations in these locations. The study sample in Chicago, IL, had significantly fewer Hispanic participants than the population in Chicago, IL. The study sample in Durham, NC, had significantly more Hispanic participants than the population in Durham, NC.

Summary

In summary, we achieved our goal to reflect the population's distribution in each of our four locations on age. The study sample was similar to the population on gender except in Chicago, IL; on African-American race except in Chicago, IL, and Sacramento, CA; and on Hispanic ethnicity except in Chicago, IL, and Durham, NC.

Exhibit 8-1a. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Chicago, IL

Demographic	Demographic Subgroup	Census Target	BCD (n = 72)	CP (n = 23)	CD (n = 66)	ODP (n = 55)	Total (n = 216)
Gender	Female	52%	65%	65%	55%	65%	62%
Age	65 years and older	10%	10%	22%	15%	11%	13%
Race	African American	36%	38%	61%	52%	40%	45%
Ethnicity	Hispanic	26%	17%	4%	18%	16%	16%

Note: Shaded cells are significantly different from Census estimates at $p < 0.05$. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 8-1b. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Durham, NC

Demographic	Demographic Subgroup	Census Target	BCD (n = 68)	CP (n = 25)	CD (n = 68)	ODP (n = 75)	Total (n = 236)
Gender	Female	52%	44%	56%	57%	51%	51%
Age	65 years and older	12%	18%	8%	12%	4%	11%
Race	African American	25%	29%	44%	28%	23%	28%
Ethnicity	Hispanic	5%	9%	12%	9%	7%	8%

Note: Shaded cells are significantly different from Census estimates at $p < 0.05$. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 8-1c. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Silver Spring, MD

Demographic	Demographic Subgroup	Census Target	BCD (n = 69)	CP (n = 28)	CD (n = 66)	ODP (n = 63)	Total (n = 226)
Gender	Female	52%	55%	57%	52%	57%	55%
Age	65 years and older	10%	20%	11%	8%	10%	12%
Race	African American	44%	33%	61%	47%	44%	44%
Ethnicity	Hispanic	9%	7%	14%	6%	8%	8%

Note: Shaded cell is significantly different from Census estimates at $p < 0.05$. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

Exhibit 8-1d. Characteristics of the study sample compared with population estimates from the U.S. Census Bureau: Sacramento, CA

Demographic	Demographic Subgroup	Census Target	BCD (n = 72)	CP (n = 21)	CD (n = 69)	ODP (n = 67)	Total (n = 229)
Gender	Female	51%	57%	52%	57%	36%	50%
Age	65 years and older	11%	15%	14%	17%	10%	14%
Race	African American	8%	13%	29%	14%	7%	13%
Ethnicity	Hispanic	16%	17%	10%	12%	18%	15%

Note: Shaded cells are significantly different from Census estimates at $p < 0.05$. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

The analyses for this chapter's research questions include subsets of the larger study population, described below.

RQ4.1 (Personal Characteristics). The analysis for RQ4.1 examining differences between demographic groups on knowledge and attitudes outcomes includes 883 of 907 participants who took part in one of the four deliberative methods and completed both the pre and post Knowledge and Attitudes Survey (97%). The surveys were administered immediately following recruitment and during the 2- to 3-week period following the assigned deliberative method, respectively.

The analysis for RQ4.1 examining differences between demographic groups on perceptions of deliberative impact includes 878 of 907 participants who took part in one of the four deliberative methods and completed the pretest Knowledge and Attitudes Survey before deliberation and the Deliberation Quality and Experience Survey at the end of their session (97%).

RQ4.2 (Concordance). The analysis of RQ4.2 on concordance included only members of the historically underrepresented group of interest.

Data Collection and Outcome Measures

Data for these analyses are drawn from the Knowledge and Attitudes pre and post surveys and the Deliberation Quality and Experience Survey. The surveys and the methods for developing outcomes are described in detail in Chapter 3.

Outcome measures. Our outcomes of interest are participants' increase in knowledge, changes in attitudes, and changes in participants' reports of the value and impact of participating in deliberation.

The knowledge items measure knowledge of issues and concepts related to health care in the United States, the use of medical evidence, and comparative effectiveness research. Seven items were scored dichotomously as 1 = correct and 0 = incorrect. (See Exhibit 3-7 for the complete list of items.) Scores are reported as the proportion correct and range from 0 to 100 percent.

To develop attitude outcomes, we used factor analysis to define the structure of survey items and determine where it was appropriate to calculate summary scores for the domains of interest. Scores are on a five-point scale (range: 1 to 5, midpoint: 3). Participants were given specific examples, or case studies, to illustrate the issues and promote deliberation. (See Chapter 2 for descriptions of case studies.) This analysis is limited to outcomes related to the hospital quality case study because all methods included this case study.

The Deliberation Quality and Experience Survey measured three overarching domains: (1) quality of communication and discourse, (2) implementation process, and (3) participant ratings of the impact of deliberation. This analysis is limited to the third domain, which reflects participants' perceptions of the effect of deliberation on participants and the value of taking part in the deliberations. Scores are on a four-point scale.

A list of attitude and experience outcomes included in this analysis is presented in Exhibit 8-2.

Exhibit 8-2. Attitude and experience outcomes

Topic	Outcome
Attitudes	
Attitudes toward using medical evidence in decisionmaking	Importance of knowing about medical evidence when making health care treatment decisions
	Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a
	Doctors and patients should consider evidence over preferences when making treatment decisions
Attitude toward mechanisms to support use of high-volume hospitals	Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a
	Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a
	Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a
Considerations of costs in decisionmaking	Doctors and patients should consider cost evidence when making treatment decisions
	People should consider the effect on group premiums when making treatment decisions ^a
Deliberation Quality and Experience	
Participant ratings of the impact of deliberation	Effect of deliberation on participants
	Perceived value of the event

^aIndicates a single item.

Modeling

Each deliberative session is a cluster of participants, and participants may have had similar outcomes because of their shared time together. It may be necessary to account for clustering by correcting standard errors caused by the intraclass correlation (ICC), or the degree to which outcomes reported by participants in the same session are correlated. We do not account for clustering of each deliberative group for knowledge and attitude outcomes because ICCs were negligible (below 0.02). We conducted multiple regression models using SAS Proc GLM for knowledge and attitude outcomes to answer RQ4.1 and RQ4.2.

In contrast, calculation of the ICCs revealed that within-session responses on the Deliberation Quality and Experience Survey were correlated. This was not unexpected, given that participants shared their group experience. We used SAS generalized estimating equations (GEE) that accounted for the clustering and adjusted standard errors for analyses of experience outcomes, thereby applying a more conservative test of significance.

RQ4.1. The equation for the models used to answer RQ4.1 is identical to the standard multiple linear regression model with the addition of the subscript “j” to indicate the jth deliberative group.

$$E_{ij} = \beta_0 + \beta_1 \text{race}_{ij} + \beta_2 \text{gender}_{ij} + \beta_3 \text{ethnicity}_{ij} + \beta_4 \text{age}_{ij} + \beta_5 \text{education}_{ij} + \delta_1 \text{income}_{ij} + \delta_2 \text{method}_{ij} + \delta_3 \text{location}_{ij} + \epsilon_{ij}$$

Where:

E_{ij} is the outcome measure of the deliberation quality and experience or the pre-post change of knowledge or attitude outcomes for participant i in deliberative group j ;

$race_i$ is an indicator that = 1 if participant i is African American and = 0 otherwise;

$gender_i$ is an indicator that = 1 if participant i is female and = 0 otherwise;

$ethnicity_i$ is an indicator that = 1 if participant i is Hispanic and = 0 otherwise;

age_i is an indicator that = 1 if participant i is age 65 or older and = 0 otherwise;

$education_i$ is an indicator that = 1 if participant i has some high school or is a high school graduate and = 0 if participant i has some college or more;

$income_i$ serves as a control variable that = 1 if participant i has a household income under \$30,000 and = 0 if \$30,000 and over;

$method_i$ serves as a control variable that is a categorical indicator of the experimental condition that participant i is assigned to (Brief Citizens' Deliberation, Online Deliberative Polling, Community Deliberation, Citizens' Panel); and

$location_i$ serves as a control variable that is a categorical indicator of the location where participant i took part in the study (Chicago, IL; Sacramento, CA; Silver Spring, MD; Durham, NC).

RQ4.2. Analyses for RQ4.2 were restricted to a specific demographic group. For example, we examined the role of concordance on outcomes among Hispanic participants only.

$$E_{ij} = \beta_0 + \beta_1 \text{concordance}_j + \delta_1 \text{race}_{ij} + \delta_2 \text{ethnicity}_{ij} + \delta_3 \text{age}_{ij} + \beta_4 \text{education}_{ij} + \delta_5 \text{income}_{ij} + \delta_6 \text{method}_{ij} + \delta_7 \text{location}_{ij} + \epsilon_{ij}$$

Where:

concordance_j is a continuous variable defined as the proportion of the participants in the index participant's deliberative session that have the same characteristic. For example, for a Hispanic person, concordance measures the proportion of participants in a deliberative group who are also Hispanic. We examined concordance for African-American participants, Hispanic participants, and participants with lower educational attainment. We did not examine concordance for age because of the relatively small percentage of elderly participants in individual groups. We examined concordance for race, ethnicity, and education because we hypothesized that participants might have different outcomes in groups concordant on these factors. Further, focus group methods often recommend gathering participants who are similar in characteristics, but our study intentionally incorporated groups with participants of different backgrounds to encourage discussions based on diverse viewpoints.²⁶

The coefficient for concordance indicates whether a higher level of group concordance was associated with better experience or larger change in knowledge or attitudes—for example, whether, for the subgroup of Hispanic participants, experience was better in groups with a higher percentage of Hispanic participants than in groups with fewer Hispanic participants. In this example, all other demographic characteristics serve as adjustment variables, allowing us to isolate the association of ethnicity while holding all other variables constant. We examined correlations among race, ethnicity, education, age, gender, and income by calculating Spearman

and Pearson correlations, and we did not find strong correlations between these variables (e.g., correlation < 0.2). All models were run in SAS Version 9.2.

Results for Research Questions on Personal Characteristics and Concordance (RQ4.1 and RQ4.2)

For clarity and ease of explanation, we present results for Research Questions 4.1 and 4.2 together.

RQ4.1 (Personal Characteristics). Do the deliberative outcomes (knowledge, attitudes, perceived impact) of African-American participants, Hispanic participants, participants age 65 and older, and participants without post-secondary education differ from the outcomes of others?

RQ4.2 (Concordance). Does group concordance, or the proportion of participants in the deliberative session who belonged to the same historically underrepresented group, affect deliberation outcomes for African-American participants, Hispanic participants, and participants without post-secondary education?

Study Population

Exhibit 8-3 shows the numbers of participants completing the Knowledge and Attitudes Survey and Deliberation Quality and Experience Survey by demographic characteristics.

Exhibit 8-3. Numbers of participants completing the Knowledge and Attitudes and Deliberation Quality and Experience Surveys, by demographic characteristics

Demographic	Demographic Subgroup	Knowledge and Attitude Outcomes, n = 883: n (Column %)	Experience Outcomes, n = 878: n (Column %)
Race	Non-African American	601 (68%)	590 (67%)
	African American	282 (32%)	288 (33%)
Gender	Male	400 (45%)	397 (45%)
	Female	483 (55%)	481 (55%)
Ethnicity	Non-Hispanic	781(88%)	779 (89%)
	Hispanic	102 (12%)	99 (11%)
Age	Under 65 years	770 (87%)	749 (81%)
	65 years and older	113 (13%)	129 (19%)
Education	Some college or more	719 (81%)	711 (85%)
	Some high school or high school graduate	164 (19%)	167 (15%)

Average concordance in a group was 31 percent for African-American participants (range, 0% to 90%), 12 percent for Hispanic participants (range, 0% to 40%), and 18 percent for participants with lower educational attainment (range, 0% to 42%).

Demographic Characteristics and Gains in Knowledge Following Deliberation

Knowledge increased from pre to post deliberation for all participants. We detected no significant associations between change in knowledge and membership in any demographic

group (race, ethnicity, age, or educational level). In addition, there was no association between concordance and knowledge among participants based on any demographic group we examined. (See Exhibits 8-4 to 8-7.)

Race: Examining Shifts in Attitudes, Deliberation Quality, and Experience

Attitudes toward using medical evidence in decisionmaking. We detected no significant associations between attitudes toward use of medical evidence in decisionmaking—including attitudes related to use of high-volume hospitals—and race, and no effect of race concordance on these outcomes among African-American participants (Exhibit 8-4).

Consideration of costs in decisionmaking. African-American participants showed less agreement and smaller changes in attitude than non-African-American participants on Factor 3, whether doctors and patients should consider cost evidence when making decisions (Exhibit 8-4).

We also found associations between race and consideration of the effect of decisions on group premiums (Item 26). African-American participants agreed more before discussions that people should consider the effects of decisions on group premiums (Item 26), but their change in attitude on this item was smaller than the change for non-African American participants (Exhibit 8-4).

We detected an association between race concordance and attitudes toward considering the effect of decisions on group insurance premiums (Item 26) among African-American participants. As the percentage of African-American participants within a group increased, agreement with this outcome also increased among African-American participants ($\beta = 1.21$, $p = 0.03$).

Participant ratings of the impact of deliberation. African-American participants had higher scores on the effect on participants and on perceived value of the event than non-African-American participants (Exhibit 8-4). Race concordance was positively associated with perceived value of the event. African Americans reported higher value when deliberating in groups with more African Americans ($\beta = 0.30$, $p = 0.02$).

Exhibit 8-4. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by race

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Knowledge	Non–African American	0.39	0.51	Reference group	-
	African American	0.37	0.47	-0.03	0.13
Attitudes Toward Using Medical Evidence in Decisionmaking					
Importance of knowing about medical evidence when making health care treatment decisions	Non–African American	4.36	4.42	Reference group	-
	African American	4.51	4.55	-0.05	0.27
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	Non–African American	2.66	2.79	Reference group	-
	African American	2.5	2.7	0.02	0.81
Doctors and patients should consider evidence over preferences when making treatment decisions	Non–African American	2.85	2.92	Reference group	-
	African American	2.99	3.04	-0.1	0.07

Exhibit 8-4. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by race (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Attitudes Toward Mechanisms To Support Use of High-Volume Hospitals					
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	Non–African American	3.11	3.28	Reference group	-
	African American	3.54	3.63	0	0.98
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	Non–African American	3.81	3.82	Reference group	-
	African American	4.01	3.96	-0.05	0.54
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	Non–African American	4.05	4.15	Reference group	-
	African American	3.99	4.09	0.02	0.8

Exhibit 8-4. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by race (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Consideration of Costs in Decisionmaking					
Doctors and patients should consider cost evidence when making decisions	Non–African American	2.6	2.83	Reference group	-
	African American	2.25	2.33	-0.22^e	0.003^e
People should consider the effect on group premiums when making treatment decisions ^a	Non–African American	2.73	2.96	Reference group	-
	African American ^c	2.83	2.9	-0.21^e	0.02^e
Participant Ratings of the Impact of Deliberation^b					
Effect of deliberation on participants	Non–African American	-	3.13	Reference group	-
	African American	-	3.29	0.16^e	< 0.0001^e
Perceived value of the event	Non–African American	-	3.56	Reference group	-
	African American ^d	-	3.7	0.12^e	< 0.0001^e

^aIndicates a single item.

^bOutcomes for “Participant ratings of the impact of deliberation” are derived from data from the Deliberation Quality and Experience Survey, which participants only took once post-deliberation. Therefore, no pre-deliberation mean scores are reported.

^cIn a separate model, we detected a positive association between race concordance and attitudes toward considering the effect of decisions on group insurance premiums among African-American participants ($\beta = 1.21$, $p = 0.03$).

^dIn a separate model, we detected a positive association between race concordance and perceived value of the event among African-American participants ($\beta = 0.30$, $p = 0.02$).

^eIndicates a significant difference at $p < 0.05$.

Note: Models hold constant ethnicity, age, education, gender, income, location, and method so that we can examine the relationship between a particular characteristic and difference in change in outcomes between groups. NA = not applicable.

Ethnicity: Examining Shifts in Attitudes, Deliberation Quality, and Experience

Attitudes toward using medical evidence in decisionmaking. Hispanic and non-Hispanic participants showed significantly different changes in attitude on considering evidence over preferences (Factor 2; $p = 0.0004$). Hispanic participants entered the study showing more agreement that doctors and patients should consider evidence over preferences than non-Hispanic participants did. Hispanic participants became more neutral and non-Hispanic participants agreed more with this proposition after discussions, with Hispanic participants exhibiting the larger shift (Exhibit 8-5). There was no effect of ethnicity concordance on any attitude toward medical evidence outcomes among Hispanic participants.

We detected no significant associations between attitudes toward mechanisms to support use of high-volume hospitals and ethnicity and no association between ethnicity concordance and these outcomes among Hispanic participants (Exhibit 8-5).

Consideration of cost in decisionmaking. We detected no significant associations between consideration of costs in decisionmaking and ethnicity and no effect of ethnicity concordance on these outcomes among Hispanic participants (Exhibit 8-5).

Participant ratings of the impact of deliberation. We detected no associations between ethnicity and effect on participants. However, Hispanic participants reported higher scores on perceived value of the event than non-Hispanic participants (Exhibit 8-5). There was no effect of ethnicity concordance on any impact of deliberation outcomes among Hispanic participants.

Exhibit 8-5. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by ethnicity

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Knowledge	Non-Hispanic	0.38	0.5	Reference group	-
	Hispanic	0.37	0.49	-0.01	0.72
Attitudes Toward Using Medical Evidence in Decisionmaking					
Importance of knowing about medical evidence when making health care treatment decisions	Non-Hispanic	4.4	4.47	Reference group	-
	Hispanic	4.43	4.46	-0.06	0.32
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	Non-Hispanic	2.57	2.75	Reference group	-
	Hispanic	2.85	2.85	-0.2	0.08
Doctors and patients should consider evidence over preferences when making treatment decisions	Non-Hispanic	2.85	2.94	Reference group	-
	Hispanic	3.18	3.04	-0.27	0.0004

Exhibit 8-5. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by ethnicity (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Attitude Toward Mechanisms To Support Use of High-Volume Hospitals					
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	Non-Hispanic	3.2	3.34	Reference group	-
	Hispanic	3.58	3.8	0.06	0.68
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	Non-Hispanic	3.85	3.84	Reference group	-
	Hispanic	4.09	4.07	-0.04	0.74
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	Non-Hispanic	4.04	4.11	Reference group	-
	Hispanic	3.96	4.25	0.18	0.06

Exhibit 8-5. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by ethnicity (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Considerations of Costs in Decisionmaking					
Doctors and patients should consider cost when making treatment decisions	Non-Hispanic	2.5	2.68	Reference group	-
	Hispanic	2.39	2.57	-0.06	0.55
People should consider the effect on group premiums when making treatment decisions ^a	Non-Hispanic	2.73	2.92	Reference group	-
	Hispanic	2.98	3.14	-0.12	0.34
Participant ratings of the impact of deliberation^b					
Effect of deliberation on participants	Non-Hispanic	NA	3.18	Reference Group	-
	Hispanic	NA	3.21	0.08	0.0523
Perceived value of the event	Non-Hispanic	NA	3.6	Reference group	-
	Hispanic	NA	3.72	0.15^c	< 0.0001^c

^a Indicates a single item.

^b Outcomes for "Participant ratings of the impact of deliberation" are derived from data from the Deliberation Quality and Experience Survey, which participants only took once post-deliberation. Therefore, no pre-deliberation mean scores are reported.

^c Indicates a significant difference at $p < 0.05$.

Note: Models hold constant race, age, education, gender, income, location, and method so that we can examine the relationship between a particular characteristic and difference in change in outcomes between groups. NA = not applicable.

Age: Examining Shifts in Attitudes, Deliberation Quality, and Experience

Attitudes toward using medical evidence in decisionmaking. We detected no significant associations between any of the outcomes related to attitudes toward using medical evidence in decisionmaking and age (Exhibit 8-6).

However, in looking at attitudes toward mechanisms to support use of high-volume hospitals, non-elderly participants' attitudes that hospitals should be required to achieve high success rates (Item 38) were similar from pre to post, but elderly participants moved from agreement toward more neutral attitudes. The difference between change from pre to post between elderly and non-elderly people on this outcome was significant in adjusted models, as shown in Exhibit 8-6.

Consideration of costs in decisionmaking. We found associations between age and consideration of the effect of decisions on group premiums (Item 26). Non-elderly participants agreed more that people should consider effects on premiums before discussions and showed more change than elderly participants (Exhibit 8-6).

Participant ratings of the impact of deliberation. We detected no associations between age and any of the impact of deliberation outcomes (Exhibit 8-6).

Exhibit 8-6. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by age

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Knowledge	Under 65 years	0.42	0.5	Reference group	-
	65 years and older	0.38	0.53	-0.02	0.38
Attitudes Toward Using Medical Evidence in Decisionmaking					
Importance of knowing about medical evidence when making health care treatment decisions	Under 65 years	4.4	4.47	Reference group	-
	65 years and older	4.43	4.41	-0.12	0.06
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	Under 65 years	2.62	2.78	Reference group	-
	65 years and older	2.48	2.62	-0.04	0.7
Doctors and patients should consider evidence over preferences when making treatment decisions	Under 65 years	2.87	2.94	Reference group	-
	65 years and older	3.01	3.02	-0.12	0.09

Exhibit 8-6. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by age (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Attitude Toward Mechanisms To Support Use of High-Volume Hospitals					
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	Under 65 years	3.25	3.41	Reference group	-
	65 years and older	3.22	3.33	-0.07	0.61
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	Under 65 years	3.89	3.91	Reference group	-
	65 years and older	3.78	3.58	-0.24	0.03
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	Under 65 years	4.01	4.13	Reference group	-
	65 years and older	4.17	4.15	-0.13	0.17

Exhibit 8-6. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by age (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Considerations of Costs in Decisionmaking					
Doctors and patients should consider cost evidence when making decisions	Under 65 years	2.44	2.64	-0.18	0.06
	65 years and older	2.8	2.88	Reference group	-
People should consider the effect on group premiums when making treatment decisions ^a	Under 65 years	2.79	2.99	-0.23	0.046
	65 years and older	2.59	2.65	Reference group	-
Participant Ratings of the Impact of Deliberation^b					
Effect of deliberation on participants	Under 65 years	NA	3.18	Reference group	-
	65 years and older	NA	3.18	0.06	0.09
Perceived value of the event	Under 65 years	NA	3.63	Reference group	-
	65 years and older	NA	3.51	-0.06	0.21

^aIndicates a single item.

^bOutcomes for “Participant ratings of the impact of deliberation” are derived from data from the Deliberation Quality and Experience Survey, which participants only took once post-deliberation. Therefore, no pre-deliberation mean scores are reported.

Note: Models hold constant race, ethnicity, education, gender, income, location, and method so that we can examine the relationship between a particular characteristic and difference in change in outcomes between groups. NA = not applicable.

Education: Examining Shifts in Attitudes, Deliberation Quality, and Experience

Attitudes toward using medical evidence in decisionmaking. We detected no significant associations between attitudes toward using medical evidence in decisionmaking and education and no effect of educational concordance on these outcomes (Exhibit 8-7).

In contrast, people with some college agreed more with disclosure of hospital volume data by doctors (Item 39) after deliberation than they had before. Attitudes among people with lower educational attainment moved from agreement to more neutral attitudes after deliberation. The difference in the change from pre to post between the two groups was significant on this outcome (Exhibit 8-7). We detected no association between educational concordance and attitudes toward mechanisms to support use of high-volume hospitals.

Consideration of costs in decisionmaking. We detected no significant associations between consideration of cost in decisionmaking and education and no effect of education concordance on these outcomes (Exhibit 8-7).

Participant ratings of the impact of deliberation. We detected no associations between education and perceived value of the event. In contrast, participants with lower educational attainment had higher scores on effect on participants than participants with higher educational attainment (Exhibit 8-7).

Exhibit 8-7. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by educational attainment

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Knowledge	Some college or more	0.38	0.51	Reference group	-
	Some HS or HS graduate	0.37	0.47	-0.03	0.12
Attitudes Toward Using Medical Evidence in Decisionmaking					
Importance of knowing about medical evidence when making health care treatment decisions	Some college or more	4.38	4.45	Reference group	-
	Some HS or HS graduate	4.53	4.52	-0.08	0.12
Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	Some college or more	2.61	2.79	Reference group	-
	Some HS or HS graduate	2.59	2.65	-0.17	0.08
Doctors and patients should consider evidence over preferences when making treatment decisions	Some college or more	2.87	2.94	Reference group	-
	Some HS or HS graduate	2.97	3.03	-0.01	0.83

Exhibit 8-7. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by educational attainment (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Attitude Toward Mechanisms To Support Use of High-Volume Hospitals					
Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	Some college or more	3.17	3.36	Reference group	-
	Some HS or HS graduate	3.57	3.55	-0.19	0.09
Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	Some college or more	3.85	3.85	Reference group	-
	Some HS or HS graduate	3.99	3.95	-0.005	0.96
Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	Some college or more	4.05	4.18	Reference group	-
	Some HS or HS graduate	3.95	3.91	-0.19^c	0.02^c

Exhibit 8-7. Unadjusted and adjusted results for knowledge, attitude, and experience outcomes, by educational attainment (continued)

Topic	Demographic Subgroup	Mean Scores at Pre-Deliberation, Unadjusted	Mean Scores at Post-Deliberation, Unadjusted	Difference in Change in Scores, Adjusted: β	Difference in Change in Scores, Adjusted: p-Value
Considerations of Costs in Decisionmaking					
Doctors and patients should consider cost evidence when making decisions	Some college or more	2.54	2.71	Reference group	-
	Some HS or HS graduate	2.27	2.51	0.07	0.4
People should consider the effect on group premiums when making treatment decisions ^a	Some college or more	2.76	2.95	Reference group	-
	Some HS or HS graduate	2.76	2.92	-0.05	0.62
Participant Ratings of the Impact of Deliberation^b					
Effect of deliberation on participants	Some college or more	-	3.15	Reference group	-
	Some HS or HS graduate	-	3.3	0.11^c	< 0.0001^c
Perceived value of the event	Some college or more	-	3.6	Reference group	-
Valued experience	Some HS or HS graduate	-	3.67	0.04	0.23

^aIndicates a single item.

^bOutcomes for “Participant ratings of the impact of deliberation” are derived from data from the Deliberation Quality and Experience Survey, which participants only took once post-deliberation. Therefore, no pre-deliberation mean scores are reported.

^cIndicates a significant difference at $p < 0.05$.

Note: Models hold constant race, ethnicity, age, gender, income, location, and method so that we can examine the relationship between a particular characteristic and difference in change in outcomes between groups. HS = high school.

Discussion

Our study did not find consistent or systematic differences between groups differing on personal characteristics in the effect of deliberation on knowledge or attitudes. More specifically, we found that demographic characteristics were not associated with changes in knowledge and were associated with changes in attitude on only selected topics. In addition, concordance was rarely associated with any outcomes. We found that the perceived impact and value of deliberation were associated with personal characteristics. These results are summarized in Exhibits 8-8 and 8-9.

The results of this analysis support four conclusions regarding deliberation and participants of different personal characteristics in this study:

- Deliberation similarly affected people from different demographic groups.
- Participants from historically underrepresented demographic groups placed more value on or perceived greater impact from their participation than others.
- Concordance (group composition) generally did not affect participant outcomes; however, there may have been an effect of concordance for African-American participants.
- Deliberation highlighted or surfaced selected content areas in which demographic groups held different views.

We elaborate on each of these results below.

Deliberation and Effects on Participants From Different Demographic Groups

This analysis offers evidence that implementing deliberation for people with a variety of backgrounds was possible and that the outcomes for participants were similar regardless of background. First, regardless of race, ethnicity, age, and educational status, participants showed similar increases in knowledge following deliberation (Exhibit 8-8). The use of educational materials combined with deliberation helped participants of many different backgrounds to understand complex and controversial topics in medical evidence, decisionmaking, and cost. Second, the direction and magnitude of the changes in attitude toward using medical evidence in decisionmaking, including mechanisms to support use of high-volume hospitals, were similar across racial, ethnic, age, and educational lines (Exhibit 8-8). Large and consistent differences between groups on knowledge and attitude outcomes would have suggested that deliberation engaged certain demographic groups more or differently; we did not observe these types of differences.

We found that participants from historically underrepresented groups—African-American and Hispanic participants and those with lower educational attainment—reported higher perceptions of the impact and/or value of deliberation. African-American and Hispanic participants reported higher value to the experience than other participants did, and African-American participants and those with lower educational attainment reported a greater impact of deliberation.

Exhibit 8-8. Summary of findings showing strength of association between knowledge, attitudes, and experience outcomes and demographic factors

Topic	Outcome	Race	Ethnicity	Age	Education
Knowledge	Knowledge of health care issues presented in educational materials	-	-	-	-
Attitudes toward using medical evidence in decisionmaking	Importance of knowing about medical evidence when making health care treatment decisions	-	-	-	-
	Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	-	-	-	-
	Doctors and patients should consider evidence over preferences when making treatment decisions	-	p = 0.0004	-	-
Attitude toward mechanisms to support use of high-volume hospitals	Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	-	-	-	-
	Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	-	-	p = 0.03	-
	Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	-	-	-	p = 0.02
Considerations of costs in decisionmaking	Doctors and patients should consider cost evidence when making treatment decisions	p = 0.003	-	-	-
	People should consider the effect on group premiums when making treatment decisions ^a	p = 0.02	-	p = 0.046	-
Participant ratings of the impact of deliberation	Effect of deliberation on participants	p < 0.0001	-	-	p < 0.0001
	Perceived value of the event	p < 0.0001	p < 0.0001	-	-

^aIndicates a single item.

Note: Blank field indicates nonsignificant finding at p < 0.05.

Group Composition in Deliberative Methods

We found that concordance—the proportion of a group represented by a specific demographic—was not associated with changes in knowledge among our participants from historically underrepresented groups (African Americans, Hispanic participants, or participants with lower educational attainment; see Exhibit 8-9). In general, we found that concordance was also not

associated with changes in attitudes about medical evidence, including use of high-volume hospitals, or with participants' perceived value or effect of deliberation.

However, we did find one exception to this result. For African-American participants, concordance (i.e., the proportion of participants in a deliberative group who are also African American) was associated with attitude change on considering the effect on group premiums when making treatment decisions (discussed below), and also with the perceived value of deliberation. Although not conclusive with regard to the impact of concordance, this finding flagged the importance of attention to group composition. In particular, African-American participants in groups with higher concordance rated the value of the event even more highly than those in lower concordant groups. African-American participants may have felt more comfortable deliberating in groups with others sharing similar background or experiences.

Exhibit 8-9. Summary of findings on strength of association between concordance and knowledge, attitudes, and experience outcomes

Topic	Outcome	Race	Ethnicity	Education
Knowledge	Knowledge of health care issues presented in educational materials	-	-	-
Attitudes toward using medical evidence in decisionmaking	Importance of knowing about medical evidence when making health care treatment decisions	-	-	-
	Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking ^a	-	-	-
	Doctors and patients should consider evidence over preferences when making treatment decisions	-	-	-
Attitude toward mechanisms to support use of high-volume hospitals	Insurance companies should require children who need heart surgery to be treated at high-volume hospitals ^a	-	-	-
	Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children ^a	-	-	-
	Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results ^a	-	-	-
Considerations of costs in decisionmaking	Doctors and patients should consider cost evidence when making treatment decisions	-	-	-
	People should consider the effect on group premiums when making decisions ^a	p = 0.03	-	-
Participant ratings of the impact of deliberation	Effect of deliberation on participants	-	-	-
	Perceived value of the event	p = 0.02	-	-

^aIndicates a single item.

Note: Blank field indicates nonsignificant finding at p < 0.05.

Highlighting or Surfacing Content Areas in Which Demographic Groups Hold Different Views

Although, as noted above, findings were consistent among participants for most outcomes, attitudes on some topic areas varied by race, ethnicity, age, or education. Because these differences were few, we concluded that they did not reflect a differential impact of deliberation as a method. However, they documented some interesting differences in views, which contribute to our substantive findings on the appropriate use of medical evidence and also highlight the importance of attention to diversity and group dynamics for designers of deliberative methods.

African American race and attitudes toward cost considerations in decisionmaking. As described earlier, we analyzed two measures of attitude addressing decisionmaking and cost:

- Doctors and patients should consider cost evidence when making treatment decisions (a factor).
- People should consider the effect on group premiums when making treatment decisions (a single item).

All participants moved from disagreement toward neutral on both of these outcomes. However, the magnitude of change on both of these measures was smaller for African-American participants than non-African-American participants at a statistically significant level—that is, although all participants moderated their views on the appropriateness of considering costs, African Americans were less inclined than others to change this view.

Reports in the literature show that race is associated with disparities in access and quality of care, and divergent experiences affect trust in and expectations of the health care system.²⁷⁻²⁹ For this reason, we believe that the difference in flexibility on views that we observed may not have been spurious, but may instead reflect differences in values and experience related to access to medical care.

An interesting caveat to this result, however, is that when a deliberative group had higher concordance on race (a larger proportion of African-American participants), the African-American participants were less likely to demonstrate this stability of view on considering the effect on group premiums when making decisions—that is, African-American participants showed greater changes in attitude. It is intriguing to consider whether groups with higher concordance offered “safer” environments, providing participants greater comfort in exploring controversial topics and greater latitude to change their views.

Attitude differences by age. As noted above, all participants moved from disagreeing to more neutral after deliberation regarding their view on the appropriateness of considering the cost impact of medical decisions. On one of the two measures we reported, elderly participants—like African-American participants—changed less than others. For elderly participants, this difference was confined to the single item measure concerning the appropriateness of considering effect of decisions on group insurance premiums.

We also observed one difference between elderly and non-elderly participants in the impact of deliberation related to the hospital quality case study. Although all participants agreed that hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children, elderly participants agreed less with this proposition than other participants—and after deliberation, they moved in the direction of still less agreement. In

contrast, non-elderly participants moved in the opposite direction, toward more agreement with this statement. The case study presented tradeoffs between high-volume hospitals many miles away versus low-volume community-based hospitals closer to home. Elderly participants may have been more sympathetic to the advantages of the local low-volume hospital, either concerning the economic benefit to the local community or the importance of easy access to the local hospital.

Attitude differences by Hispanic ethnicity. Hispanic participants agreed more than others before deliberation that providers and patients should consider evidence over preferences when making treatment decisions. Following deliberation, Hispanic participants' views moderated and their scores drew closer to those of non-Hispanic participants, but they continued to show more support for consideration of evidence over preferences.

Differences by educational level. People with higher versus lower educational attainment moved in opposite directions with regard to agreement with the statement that doctors should tell patients about high-volume hospitals with better outcomes. Participants may have been expressing values about the role of physicians through this outcome. For example, some participants expressed the view that physicians are primarily business people and were less convinced of physicians' role in providing this type of information. We did not track these differences in views by participants' personal characteristics, but differing views on the role of physicians, if associated with educational level, may underlie the result we observed.

Conclusion

Our results concerning the effect of deliberation on participants with differing personal characteristics indicated that, overall, participants experienced very similar changes in knowledge and attitudes and were similarly engaged by the process. The primary differences we observed with respect to the perceived impact of deliberation on historically underrepresented demographic groups were positive. African-American and Hispanic participants reported a higher value to the experience than other participants, and African-American participants and those with lower educational attainment reported a greater impact of deliberation.

We also found little evidence that group composition, concordance, had an impact on shifts in knowledge or attitude. However, the impact of concordance is an area that merits additional study. Our findings indicated that African-American participants in groups with higher concordance valued the deliberative events more highly than African-American participants in groups with lower concordance. We also found one difference in attitude change for African-American groups with higher concordance, suggesting that, at a minimum, group composition must be considered carefully by those designing deliberative sessions.

Chapter 9. Research Question 5: Group Behavior—Consensus, Polarization, and Domination

In this chapter, we discuss outcomes related to Research Question 5: Do the group-level effects (i.e., the internal group dynamics) of public deliberation vary by deliberative method? Public deliberation is intended to create an environment in which members of the public reason with each other as equals. If some members of a deliberative group are unduly influenced by others with greater social or economic status, then the full range of opinion within the group might not be reflected accurately, which would violate the principles of deliberative democracy. Thus, it is important to understand whether group members are arriving at their conclusions based on the substance of competing arguments, or whether their conclusions are influenced by artifacts of group psychology.

Critics of deliberation claim that two kinds of bias-inducing behavior are inherent in deliberative groups: polarization and domination.^{30,31} Evidence for these criticisms comes primarily from research on the behavior of juries, where jurors are compelled to reach a decision based on unanimity or consensus. Studies of deliberation modeled on juries have produced a literature on the “law of group polarization,” a pattern whereby groups regularly move away from the midpoint toward more extreme views, regardless of the content of the deliberation.³¹ Jury research has also revealed domination by people who are traditionally more privileged in society, such as men, whites, and people who are better educated. Therefore, critics argue that deliberation, especially in the form of juries or jury-like processes, produces conclusions distorted by polarization and societal inequalities.

As noted in Chapter 2, consensus was not a goal in the Community Forum Deliberative Methods Demonstration. Consensus was not discouraged if it occurred, but there was no intent to achieve it. The goal of consensus distinguishes jury deliberation from the Community Forum methods and might explain why the inequitable patterns of group behavior observed in juries might not be found in other forms of deliberation. To explore these issues, we identified the following research questions (RQs):

RQ5. Do the group-level effects (i.e., the internal group dynamics) of public deliberation vary by deliberative method?

RQ5.1. Were any deliberative methods more likely to reach consensus than others?

RQ5.2. Does deliberation tend to polarize participants’ views?

RQ5.2.1. Does polarization occur?

RQ5.2.2. If polarization occurs, does it vary by deliberative method?

RQ5.2.3. If polarization occurs, does it vary by measure?

RQ5.3. Do members of the group from traditionally privileged demographic groups dominate deliberation?

RQ5.3.1. Does domination occur?

RQ5.3.2. If domination occurs, does it vary by deliberative method?

RQ5.3.3. If domination occurs, does it vary by measure?

Although comparisons among methods are an important part of the following discussion, the results for Citizens' Panel (CP) at the group level are constrained by the small number of CP groups that were conducted—4 compared with 24 for each of the other methods—which makes the estimates for CP inherently less stable at the group level.

Data Sources and Methods

These analyses use data from the pre and post Knowledge and Attitudes Survey. In this analysis, the outcome measures are participants' individual item responses and composites, scored for each deliberative group by averaging the scores for each participant in the group. The group score—the average for participants in each session—for each measure was computed separately for the pre and post surveys so that we could calculate change at the group level.

As explained in Chapter 3, we included several outcomes in our analyses that are relevant across all deliberative methods and allow for interpretation across all the RQs. However, the items and composites used as outcome measures for constructing these group-level scores differ slightly in some cases from those used in the participant-level analyses in Chapters 5–8 (RQs 1–4). We defined some composite measures differently at the group level, because the definition used at the individual level was not as meaningful for groups. For example, at the group level analysis, Factor 2 (*Doctors and patients should consider evidence over preferences when making treatment decisions*) was separated into Factor 2a (*Doctors should consider evidence over preferences when making treatment decisions*) and Factor 2b (*Patients should consider evidence over preferences when making treatment decisions*). We did this because we wanted to break out group-level differences on views of doctors versus patients with respect to the use of evidence over preferences in making treatment decisions. In some cases, we removed an item from a composite and treated it as its own outcome at the group level, because we were particularly interested in the content of the item.

Of the 16 outcomes used for group-level analysis in this chapter, 8 were identical to those used in the individual-level analyses in Chapters 5–8 and 8 were redefined. Exhibit 9-1 (Exhibits 9-1a through 9-1d) defines the eight measures used in the group-level analyses that were not used for individual-level analysis in Chapters 5–8. The definitions of the eight group-level measures that were also used in the individual-level analyses may be found in Chapter 3, Exhibit 3-7.

The results presented for polarization and domination are based on the direction of movement in group scores, without regard to the size of the shift. Small shifts and large shifts are given equal weight in the analysis. If small shifts were to be given less weight or discounted entirely, we would be less likely to find polarization or domination patterns. By weighting shifts equally, regardless of size, we have made it easier to find polarization and domination.

Exhibit 9-1a. Measures used for groups that were not used for individual-level analyses: government

Measure #	Composites
KA Q30	The government should limit advertising of unhealthy food to children.
KA Q31	The government should tax unhealthy food to discourage people from buying it.
KA Q32	The government should limit the sale of unhealthy foods in the United States.
KA Q33	The government should ban the sale of food that research has shown to cause heart disease.
KA Q36	The government should limit when doctors can prescribe antibiotics.
KA Q40	People who are obese have more health problems and live shorter lives than people with healthy weights. How much do you support the government spending money to help obese people lose weight?
KA Q41	How much do you support the government spending money to help people stop smoking?

Exhibit 9-1b. Measures used for groups that were not used for individual-level analyses: Factor 2a – Doctors should consider evidence over preferences when making treatment decisions

Measure #	Composites
KA Q15	Doctors should be able to provide a medical treatment even if medical research results show it won't work for the patient.
KA Q16	Doctors should be able to provide a medical treatment even if medical research results show it might not work for the patient.
KA Q17	Doctors should be able to provide a medical treatment they prefer, even if medical research shows that there is another treatment that works better.

Exhibit 9-1c. Measures used for groups that were not used for individual-level analyses: Factor 2b – Patients should consider evidence over preferences when making treatment decisions

Measure #	Composites
KA Q18	People should be able to get a medical treatment they prefer, even if medical research results show it won't work for them.
KA Q19	People should be able to get a medical treatment they prefer, even if medical research results show it might not work for them.
KA Q20	People should be able to get a medical treatment they prefer, even if medical research shows there is another treatment that works better.
KA Q34	People should be able to get an antibiotic if they want it, even if it might not help.

Exhibit 9-1d. Measures used for groups that were not used for individual-level analyses: individual indexes

Measure #	Individual Indexes
KA Q13	In your opinion, which of the following best determines when medical treatments are "proven to be effective": 1 = by the results of medical research. 2 = by each doctor based on his or her own experience. 3 = by people who know their bodies best and can say what is effective and what is not.
KA Q25	Health insurance should pay for a treatment that a person prefers, even if medical research results show that it does not work.
KA Q27	Parents should limit how much unhealthy food their children eat.
KA Q28	Schools should limit how much unhealthy food is served to children.
KA Q29	Teachers should teach their students healthy eating habits.

Consensus

We determined if groups achieved consensus by computing the pre-deliberation (Time 1) and post-deliberation (Time 2) standard deviations of participant scores for each of the 76 groups on three selected outcome measures:

- Factor 1: Importance of knowing about medical evidence when making health care treatment decisions
- Factor 2: Providers and patients should consider evidence over preferences when making treatment decisions
- Item 12: Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking

We chose these 3 out of the total 16 outcome measures used for group-level analyses to evaluate consensus because they most directly address the central tradeoff between the use of evidence and individual freedom when making medical decisions. We defined consensus as a shift from a larger standard deviation at Time 1 to a smaller standard deviation at Time 2. Thus, for each group, we subtracted the pre-intervention standard deviation from the post-intervention standard deviation, so that a negative result indicated a move toward consensus and a positive result indicated a move away from consensus following deliberation. If consensus were a random process, we would expect about 50 percent of groups to have negative signs on the change in standard deviation; if consensus were systematic, we would expect the percentage to approach 100 percent. Exhibit 9-2 provides the number and percent of groups moving toward consensus on all three measures for the 76 groups by deliberative method.

Exhibit 9-2. Number and percent of groups moving toward consensus by method and measure

Deliberative Method	Consensus Score ^a Factor 1: Importance of knowing about medical evidence when making health care treatment decisions	Consensus Score ^a Factor 2: Providers and patients should consider evidence over preferences when making treatment decisions	Consensus Score ^a Item 12: Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking
BCD (n = 24)	12 50%	9 38%	10 42%
CP (n = 4)	2 50%	2 50%	2 50%
CD (n = 24)	13 54%	14 58%	11 46%
ODP (n = 24)	16 67%	9 38%	13 54%
All methods and groups (n=76)	43 57%	33 43%	35 46%

^aStandard deviation Time 2 – standard deviation Time 1.

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

For all three measures, only about half the groups moved toward consensus following deliberation, which suggests that achieving consensus was a random process. Online Deliberative Polling (ODP) was the only method with a consensus indicator exceeding 60

percent and then only for Factor 1 (*Importance of knowing about medical evidence when making health care treatment decisions*).

Exhibit 9-3 displays the mean change in standard deviation for each of the three measures by deliberative method. The mean takes into account the magnitude of the change as well as the direction.

Exhibit 9-3. Mean change in standard deviation by deliberative method and measure^a

Method	Mean Change in Standard Deviation Factor 1: Importance of knowing about medical evidence when making health care treatment decisions	Mean Change in Standard Deviation Factor 2: Providers and patients should consider evidence over preferences when making treatment decisions	Mean Change in Standard Deviation Item 12: Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking
BCD	0.02	0.05	0.08
CP	-0.02	0.00	0.02
CD	-0.05	-0.05	0.02
ODP	-0.09	-0.02	0.00
Overall	-0.04	-0.01	0.03

^aA negative mean change indicates a move toward consensus following deliberation, and a positive mean change indicates a move away from consensus following deliberation.

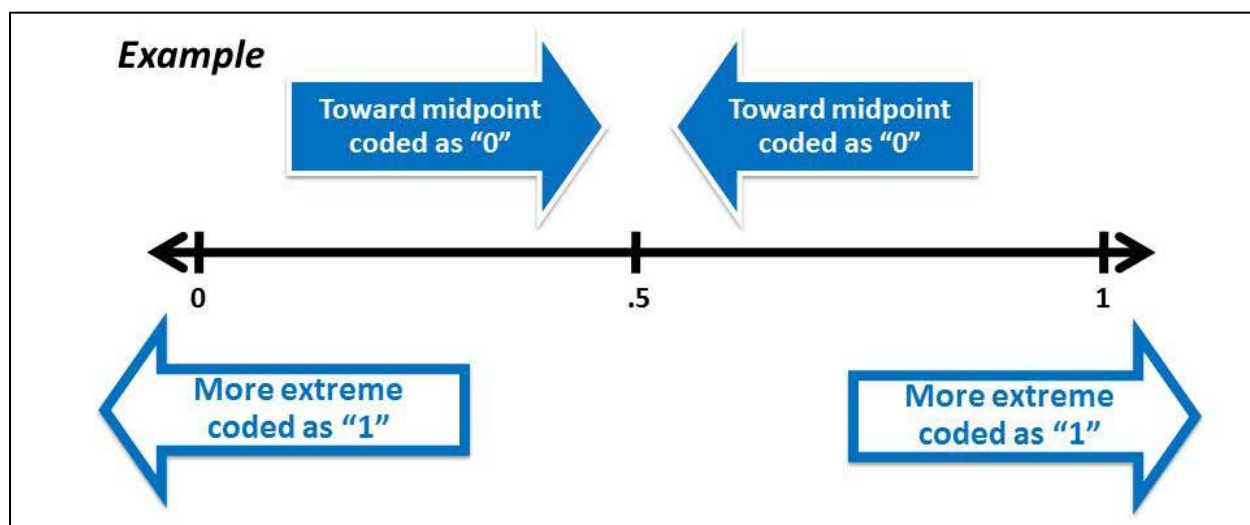
Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; ODP = Online Deliberative Polling.

On average, Brief Citizens' Deliberation (BCD) groups did not move toward consensus on any of the measures, and none of the methods moved toward consensus on Item 12 (*Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking*). Overall, there was little systematic movement toward consensus in the Community Forum groups, and none of the methods systematically reached consensus on any of these three measures.

Polarization

Group polarization in public deliberation is the systematic tendency of groups and the individuals who compose them to move "toward a more extreme point in whatever direction is indicated by the members' pre-deliberation tendency."¹³ In other words, groups that on average agree with a policy or assertion will tend to agree with it more strongly following deliberation, and groups that disagree with it will tend to disagree with it more strongly following deliberation as a result of an inherent characteristic of group behavior that is independent of the content of the deliberation. Exhibit 9-4 illustrates this behavior.

Exhibit 9-4. Illustration of polarization



The unit of analysis for evaluating polarization is the opportunity for polarization to occur. The number of opportunities is the product of the number of groups multiplied by the number of measures computed for each group. In this analysis, each of the 16 outcome measures was computed for each of the 76 deliberative groups, which yields a grand total of 1,216 opportunities for groups to have moved either toward or away from the extreme positions. The percentage of opportunities in which the group moved toward the extremes, which we call the opportunity score, serves as the indicator of polarization.

If movement toward the extremes (i.e., polarization) is an inherent, inevitable characteristic of group deliberation, then a high percentage of groups should move toward the extremes following deliberation, regardless of content or deliberative method. We would not expect all groups to move toward the extremes in the presence of polarization, because there will be some random exceptions. If the Community Forum Deliberative Methods Demonstration was free of polarization, then we would expect about 50 percent of the opportunities to move toward the extremes and about 50 percent to move toward the midpoint. Each group has two alternatives on each measure—move toward the extreme or toward the middle—so half the opportunities moving in each direction would indicate random variation and the absence of polarization.

Did Polarization Occur?

Using all 1,216 observations, or opportunities for attitudes to move toward the extremes following deliberation, the top right cell of Exhibit 9-5 indicates that 45 percent of these observations demonstrated movement toward the extremes. Because this rate or opportunity score is close to 50 percent, it implies that movement toward the extremes occurred randomly and is not systematic or inevitable.

The top row of Exhibit 9-5 also indicates that the all-measures-combined opportunity score for each deliberative method was close to 50 percent. ODP had the lowest opportunity score at 43 percent, followed by Community Deliberation (CD) at 45 percent, BCD at 46 percent, and CP at 55 percent. Looking at the individual measures within each deliberative method, the highest opportunity score for ODP was 67 percent (Item 13: *What determines when medical treatments*

are “*proven to be effective*”); for CP, 100 percent (Factor 1: *Importance of knowing about medical evidence when making health care treatment decisions*); for BCD, 85 percent (Factor 4: *Role of community [parents, teachers, and schools] in limiting unhealthy food*); and for CD, 83 percent (Item 13: *What best determines when medical treatments are “proven to be effective”*). Thus, although a few measures display high percentages of movement to the extremes, there is no pattern of polarization among any of the methods.

Did Polarization Vary by Measure?

When honing in on each individual measure combined over the four methods (the right-hand column of Exhibit 9-5), only 5 of the 16 measures exceeded 50 percent and only 1 measure exceeded 70 percent (Item 13: *What best determines when medical treatments are “proven to be effective”*). To illustrate, the opportunity scores on Factor 3 (*Doctors and patients should consider cost evidence when making treatment decisions*) range from 0 percent to 60 percent; only BCD exceeds 50 percent. On Factor 2b (*Patients should consider evidence over preferences when making treatment decisions*), all four of the CP groups moved to the extremes, while the opportunity scores for the other methods and the average over all the methods were well below 50 percent. Factor 4 (*Role of community [parents, teachers, and schools] in limiting unhealthy food*), Item 13 (*What best determines when medical treatments are “proven to be effective”*), and Item 25 (*Health insurance should pay for a treatment that a person prefers, even if medical research results show that it does not work*) were the only measures for which at least three methods displayed opportunity scores above 50 percent; Item 13 was the only measure for which all methods exceeded 50 percent. Nine of the 16 measures had ranges over the four deliberative methods that spanned 50 percent, while the remaining 7 measures had ranges that were all below 50 percent. Only one measure—Item 13—had opportunity scores that were consistently high compared with the other measures (between 67% and 100% of the groups polarizing for all four deliberative methods). Thus, with the possible exception of Item 13, there is no evidence that some measures are more susceptible to polarization than others.

Exhibit 9-5. Percentage of group-by-measure opportunities moving toward the extremes, by deliberative method

Measure	Community Deliberation (n = 24 Groups)	Brief Citizens' Deliberation (n = 24 Groups)	Citizens' Panel (n = 4 Groups)	Online Deliberative Polling® (n = 24 Groups)	All Deliberative Methods (n = 76 Groups)
All measures combined	45%	46%	55%	43%	45%
Factor 1. Importance of knowing about medical evidence when making health care treatment decisions	67%	46%	100%	54%	58%
Factor 2. Doctors <i>and</i> patients should consider evidence over preferences when making treatment decisions.	29%	42%	50%	25%	33%
Factor 2a. Doctors should consider evidence over preferences when making treatment decisions.	42%	50%	75%	33%	43%
Factor 2b. Patients should consider evidence over preferences when making treatment decisions.	33%	33%	100%	38%	38%
Factor 3. Doctors and patients should consider cost evidence when making treatment decisions.	25%	60%	0%	33%	37%
Factor 4. Role of community (parents, teachers, and schools) in limiting unhealthy food.	58%	85%	75%	38%	61%
Factor 5. Role of government policies in limiting unhealthy food.	46%	50%	25%	42%	45%
Government	33%	42%	50%	46%	41%
Item 12. Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking.	17%	29%	0%	50%	30%

Exhibit 9-5. Percentage of group-by-measure opportunities moving toward the extremes, by deliberative method (continued)

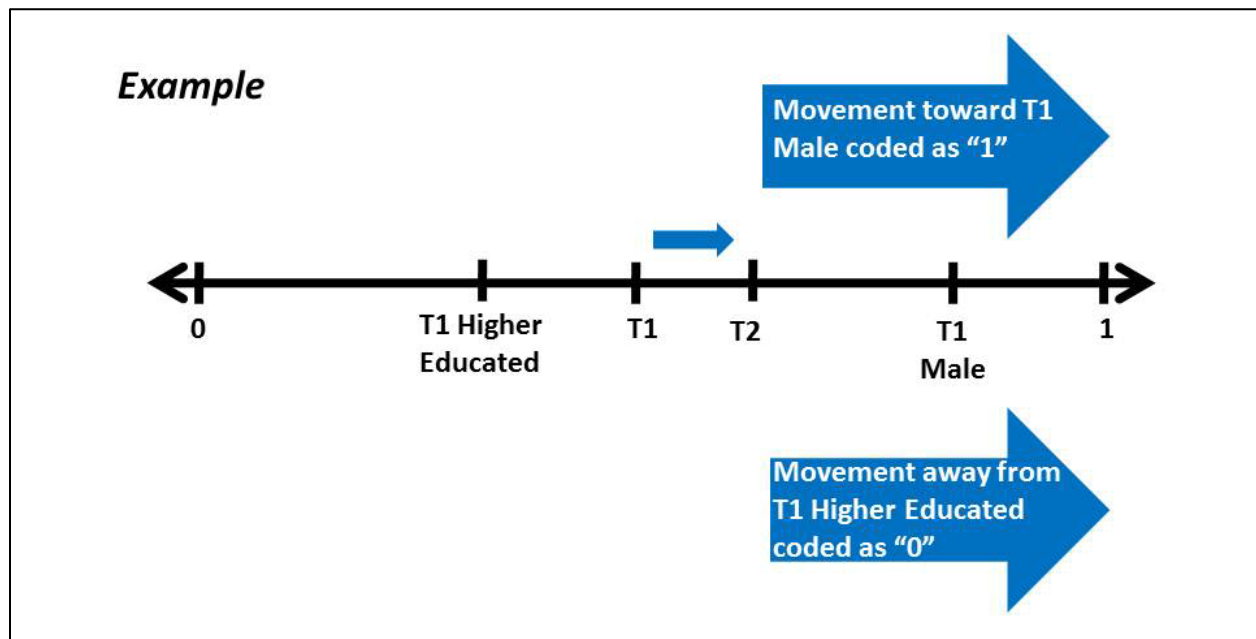
	Community Deliberation (CD) (n=24 groups)	Brief Citizens' Deliberation (BCD) (n=24 groups)	Citizens' Panel (CP) (n=4 groups)	Online Deliberative Polling® (ODP) (n=24 groups)	All Deliberative Methods (n=76 groups)
Item 13. What best determines when medical treatments are "proven to be effective."	83%	71%	100%	67%	75%
Item 25. Health insurance should pay for a treatment that a person prefers, even if medical research results show that it does not work.	67%	50%	75%	54%	58%
Item 26. People should consider the effect on group premiums when making treatment decisions.	21%	38%	25%	29%	29%
Item 27. Parents should limit how much unhealthy food their children eat.	46%	33%	25%	54%	43%
Item 28. Schools should limit how much unhealthy food is served to children.	46%	13%	75%	33%	33%
Item 29. Teachers should teach their students healthy eating habits.	67%	46%	75%	46%	54%
Item 37. Insurance companies should require children who need heart surgery to be treated at high-volume hospitals.	42%	42%	25%	46%	42%

Domination

Domination occurs when the group routinely adopts the opinions of people who are traditionally more privileged in society—such as men, whites, and people who are better educated—as its predominant conclusions.³⁰ This analysis uses the same basic approach as the polarization analysis—opportunity scores computed from the group-by-measure combinations—except that the criterion is movement by the group *following deliberation* toward the *pre-deliberation* opinions of the privileged members of the group. Exhibit 9-6 illustrates this concept. Time 1 and Time 2 indicate the pre and post deliberation means of one entire deliberative group on one attitudinal measure. Time 1 Male and Time 1 Higher Educated indicate the pre-deliberation means of two privileged subsets of the group's membership. The group's mean shifted to the right on the scale (Time 1 → Time 2), which is toward the pre-deliberation mean of the men in the group. This shift meets the definition of domination. At the same time, the group mean shifted away from the mean of its better educated members, which is an opportunity in which domination did not occur. Exhibit 9-6 illustrates two opportunities for domination by privilege, only one of which was realized.

Like the polarization model described above, the combination of 76 deliberative groups by 16 outcome measures yields 1,216 opportunities for domination. However, we also use five indicators of privilege in the domination analysis: gender (being male), marital status (being married), education (having a 4-year college degree or more education), household income (\$60,000 or more per year), and race (identifying as white). Therefore, the grand total number of opportunities for domination is 6,080. The proportion of opportunities in which domination occurred is the opportunity score, or the indicator of domination, across multiple opportunities.

Exhibit 9-6. Illustration of domination



Note: T = Time.

As with polarization, if domination is an inherent, inevitable characteristic of deliberation, we would expect almost all groups to exhibit it, and the opportunity score would be close to 100 percent. If domination is not inevitable, we would expect the opportunity score to be close to 50 percent, or a random distribution between dominated and not dominated.

Exhibit 9-7 lists the five indicators of social privilege and the coding used to define privilege in the domination analyses.

Exhibit 9-7. Definition of social privilege items

Social Privilege Variable	Survey Question	Not Privileged	Privileged
Male	Are you male or female?	0 = Female	1 = Male
Married	What is your current marital status?	0 = Other	1 = Married
Educated	What is the highest grade or level of school that you have completed?	0 = Not completed college	1 = 4-year degree and above
High income	Using the categories below, please indicate the annual income of your household	0 = Below \$60,000	1 = \$60,000 or more
White	What is your race?	0 = Not white	1 = White

For the domination analyses, we calculated the opportunity scores, or the percentage of the opportunities in which a group at Time 2 (after deliberation) moved toward the Time 1 (before deliberation) position of its more privileged members, for each of the 16 outcome measures.

The analysis was conducted separately for each of the five indicators of privilege listed in Exhibit 9-7. For instance, we determined the percentage of groups that, at Time 2, scored closer to the Time 1 position of its male members. We then computed the same percentages using the Time 1 average position of the group's married members, and so on for the remaining three indicators of privilege described in Exhibit 9-7. This approach produces a large number of opportunity scores, 6,080, a number too cumbersome to present in this report. Thus, we averaged the opportunity scores for each individual indicator of privilege, stratified by deliberative method and outcome measure. Exhibit 9-8 presents these opportunity scores averaged over the five privilege indicators for each of the 16 outcome measures. If domination were systematic and inevitable, most of the opportunity scores would be close to 100 percent. If domination were not systematic, most of the opportunity scores would cluster around 50 percent.

Did Domination Occur?

If domination occurred systematically in our deliberative groups, the majority of opportunity scores for domination would be close to 100 percent. However, the right-hand column in Exhibit 9-8 shows that the vast majority of percentages clustered around 50 percent, indicating that domination by privileged group members occurred randomly and not systematically.

Did Domination Vary by Deliberative Method and Measure?

We did not find evidence of domination within any of the four deliberative methods (top row of Exhibit 9-8) or within any of the 16 outcome measures (right-hand column of Exhibit 9-8). However, there were seven instances across all methods and outcome measures where percentages reached or exceeded 60 percent; of these, five were in the CP method (Exhibit 9-8). There are several possible interpretations for this result. One interpretation is that the CP method might be more susceptible to domination than the other methods, even if there is no evidence of broad-based domination. It is possible that the long period of deliberation used by the CP method, 2.5 days, provides privileged people with more time to sway other members of the group. Alternatively, better educated participants might have grasped the complex material more quickly than less educated participants, and lengthier deliberation enabled better educated participants to convey that understanding to other group members. It is also possible that the small number of CP groups—4 compared with 24 for each of the other methods—makes the CP opportunity scores inherently more variable than the opportunity scores for the other three methods.

Exhibit 9-8. Percentage of groups that moved at Time 2 toward the Time 1 position of privileged members

Measure	Community Deliberation (24 Groups)	Brief Citizens' Deliberation (24 Groups)	Citizens' Panel (4 Groups)	Online Deliberative Polling (24 Groups)	All Methods (76 groups)
All measures combined	46%	51%	51%	53%	50%
Factor 1. Importance of knowing about medical evidence when making health care treatment decisions.	46%	47%	45%	53%	48%
Factor 2. Doctors and patients should consider evidence over preferences when making treatment decisions.	46%	53%	35%	54%	50%
Factor 2a. Doctors should consider evidence over preferences when making treatment decisions.	50%	48%	50%	54%	51%
Factor 2b. Patients should consider evidence over preferences when making treatment decisions.	46%	48%	20%	54%	48%
Factor 3. Doctors and patients should consider cost evidence when making treatment decisions.	53%	53%	90%	63%	58%
Factor 4. Role of community (parents, teachers, and schools) in limiting unhealthy food.	43%	53%	55%	56%	51%
Factor 5. Role of government policies in limiting unhealthy food.	48%	53%	40%	59%	53%
Government	34%	41%	45%	59%	45%
Item 12. Medical research versus doctor's knowledge about patient as most important in medical treatment decisionmaking.	53%	41%	75%	56%	51%
Item 13. What best determines when medical treatments are "proven to be effective."	50%	52%	70%	53%	53%
Item 25. Health insurance should pay for a treatment that a person prefers, even if medical research results show that it does not work.	52%	58%	35%	48%	52%
Item 26. People should consider the effect on group premiums when making treatment decisions.	40%	47%	60%	50%	46%
Item 27. Parents should limit how much unhealthy food their children eat.	39%	54%	65%	48%	48%
Item 28. Schools should limit how much unhealthy food is served to children.	45%	64%	50%	52%	53%
Item 29. Teachers should teach their students healthy eating habits.	47%	47%	30%	42%	45%
Item 37. Insurance companies should require children who need heart surgery to be treated at high-volume hospitals.	48%	58%	55%	54%	53%

Note: Privilege is defined as being male, white, married, high education, or high income.

Discussion

The deliberative methods employed by the Community Forum Deliberative Methods Demonstration did not attempt to achieve consensus. Although group members shared their opinions with each other, the results presented here are based on each participant's reports in

confidential questionnaires outside the group sessions, which distinguish all four methods from juries. Furthermore, they are all designed to engage participants in sharing reasons with a balanced agenda of competing arguments.

The picture that emerges from these analyses is that small-group distortions found in jury-like settings are not evident in these data. We do not find the pattern of polarization or systematic movement toward more extreme positions, nor do we find any systematic pattern of domination by the more privileged members of society. Lastly, we do not find a systematic pattern of movement toward consensus. While consensus may seem desirable, the social pressure of false consensus would not be. If we found consensus systematically in the confidential questionnaires, such a pattern of small-group psychology would also raise questions as to whether “group think” was determining the conclusions, rather than conclusions being determined by the substance of issues and the values that each participant brought to the deliberation.

Given the careful preparations to have balanced educational materials for all groups, it is reassuring that we do not find any of these three patterns—polarization, domination, or consensus—systematically in the Community Forum deliberations. The results found here may offer an argument for designing deliberative consultations with these design features in mind: no shared consensus seeking; confidential questionnaires; and well-tested, high-quality, and balanced educational materials.

Chapter 10. Evaluation of Costs

In this section we describe the costs associated with implementing deliberative sessions. Cost can be an important consideration for a sponsor in choosing among approaches to deliberation. To assist with this calculation, this chapter presents estimates of implementation costs for each deliberative method.

Costs Included and Excluded—and Why

The main costs of deliberation are associated with developing materials, recruiting participants, holding sessions, and analyzing and reporting results. In this analysis, we focus on one of these categories, presenting costs associated with holding the deliberative sessions. The costs of developing materials can vary dramatically, depending on the topic and the knowledge required for informed discussion. For this reason, we do not present costs associated with having experts participate in the deliberative sessions or with developing, producing, or distributing educational materials, case studies, surveys, or other materials used in deliberative sessions. Costs associated with analysis and reporting results vary similarly, depending on the type of report desired by the sponsor.

In addition, we exclude from this presentation the research-specific costs we incurred, which a sponsor holding deliberative sessions outside the research context would not incur. For instance, the oversight and adaptations needed to ensure recruitment and random assignment of a diverse sample to support our analyses would not be required for a typical implementation of deliberative discussions.

We also exclude fixed costs that do not vary across the discussion methods. For example, we created a Web site through which participants in all deliberative methods accessed educational materials; we excluded costs associated with developing this site. In addition to Web development costs, we excluded fixed costs associated with planning, our own travel, and educational materials development. We also excluded overhead costs, since other organizations will have different indirect cost structures from those of the American Institutes for Research.

Finally, we excluded facilitation costs from this analysis. In our demonstration, we drew upon the skills of expert deliberative facilitators for many of our sessions. These individuals were consultants to our project and also provided extensive training for members of our regular staff and new Online Deliberative Polling (ODP) moderators. We did not track these costs separately and for that reason do not present them here.

We present actual incurred costs of participant incentives, facilities, equipment, and supplies needed for deliberative groups, broken down by type of method. Exhibit 10-1 shows the number of participants and hours of deliberations by method, while Exhibit 10-2 shows the cost components, units, and costs.

Exhibit 10-1. Number of participants and total hours of deliberation, by method

Method	Number of Participants	Total Hours of Deliberation
Citizens' Panel	24	20
Brief Citizens' Deliberation	12	2
Community Deliberation	12	6
Online Deliberative Polling	12	5

Exhibit 10-2. What costs were associated with holding deliberative sessions?

Cost Components	Unit	BCD	CD	CP	ODP	Unit Amount
Participant Costs						
Incentive	Per person	12 people	12 people	24 people	12 people	Varied ^a
Childcare	Per hour	2 hours	5 hours	20 hours	NA	\$10.00
Transportation	Per person per day	12 people * 1 day	12 people * 2 days	24 people * 3 days	NA	\$10.00
Facilities						
Site rental in first session	Per room per day	1 room * 0.5 day	1 room * 0.5 day	3 rooms * 1 day	NA	\$367.67
Site rental in second session	Per room per day	NA	1 room * 0.5 day	3 rooms * 1 day	NA	\$367.67
Site rental in third session	Per room per day	NA	NA	1 room * 0.75 day	NA	\$367.67
Food and drink – participant breakfast	Per person per day	NA	NA	24 people* 3 days	NA	\$16.00
Food and drink – participant lunch	Per person per day	12 people * 1 day	12 people * 2 days	24 people* 3 days	NA	\$22.65
Food and drink – participant snack	Per person per day	NA	NA	24 people* 1 day	NA	\$9.00
Food and drink – participant snack	Per person per day	NA	NA	24 people* 1 day	NA	\$24.01

Exhibit 10-2. What costs were associated with holding deliberative sessions? (continued)

	Unit	BCD	CD	CP	ODP	Unit Amount
Equipment, Technology						
Microphones and speakers	Per day	NA	NA	2 days	NA	\$205.00
Projectors and screens in first session	Per room	1 room	1 room	3 rooms	NA	\$387.50
Projectors and screens in second session	Per room	NA	1 room	3 rooms	NA	\$387.50
Projectors and screens in third session	Per room	NA	NA	1 room	NA	\$387.50
Computers	Per computer	3 computers	3 computers	3 computers	3 computers	\$699.00
Hard-line Internet connection in hotel	Per day	NA	NA	2 days	NA	\$233.33
Office-based Internet access	Per week	NA	1 week	NA	1 week	\$600.00
Telephone/conference line	Per group	NA	NA	NA	1 group	\$375.00
WebEx	Per license	NA	NA	NA	1 license	\$312.00
Supplies						
Pens	Per pen per person	1 pen * 12 people	1 pen * 12 people	1 pen * 24 people	NA	\$0.12
Paper and copies	Per copy	24 copies * 12 people	35 copies * 12 people	111 copies * 24 people	NA	\$0.35
Flipcharts	Per chart	1 chart	2 charts	11 charts	NA	\$50.00
Easels	Per easel	1 easel	1 easel	5 easels	NA	\$25.00
Markers	Per marker box	4 marker boxes	4 marker boxes	20 marker boxes	NA	\$4.00

^aVaried depending on the method. Participants in BCD received \$100; CD received \$160; CP received \$325; and ODP received \$125.

Note: BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; NA = not applicable; ODP = Online Deliberative Polling.

Analysis of Costs

We present costs for each method separately and calculate the ratio of costs of each method compared with the least costly method, Brief Citizens' Deliberation (BCD). To provide a sensitivity analysis, we estimated low and high rates for all participant and facility inputs, as well as selected equipment, technology, and materials inputs (microphones and speakers, projectors and screens, computers, and copies). Low estimates for participant inputs were set at \$0 to provide an estimate when no reimbursement or incentive is provided to participants. High estimates reflect a reimbursement of \$15 for childcare and transportation costs, based on costs in Sacramento, CA. The low and high estimates are based on costs at certain locations for the facility, equipment, technology, and material inputs. For example, the low rate for site rental was based on the cost in Chicago, IL, while the high rate for site rental was based on Silver Spring, MD. For certain items not expected to show significant variation, such as basic office supplies, we used the base value instead of considering low or high alternative estimates.

Results

Costs are presented in Exhibit 10-3. The least costly method to implement was BCD (\$4,500), while the most costly method to implement was Citizens' Panel (CP; \$23,500). The cost for our online method, ODP, was very close to BCD at \$4,900, and Community Deliberation (CD) was somewhat more costly at \$6,900. Compared with the least costly method, BCD, ODP was 1.1 times higher, CD 1.5 times higher, and CP 5.3 times higher in cost.

The largest area of cost that we tracked was, for BCD, CD, and ODP, that of equipment and technology, accounting for more than half of costs. In contrast, the greatest area of cost for CP was participant-related costs.

Exhibit 10-3. Costs per deliberative session: base estimate, low- and high-range estimates, by method (2012 U.S. dollars)

Cost Component	BCD: n = 12; 1 Day Base (Low; High)	CP: n = 24; 3 Days Base (Low; High)	CD: n = 12; 2 Days Base (Low; High)	ODP: n = 12; No Onsite Meeting Base (Low; High)
Participants				
Incentive	\$1,200 (\$0; \$1,200)	\$7,800 (\$0; \$7,800)	\$1,920 (\$0; \$1,920)	\$1,500 (\$0; \$1,500)
Childcare	\$20 (\$0; \$30)	\$1,565 (\$0; \$2,347)	\$87 (\$0; \$130)	NA
Transportation	\$120 (\$0; \$180)	\$720 (\$0; \$1,080)	\$240 (\$0; \$360)	NA
Subtotal	\$1,340 (\$0; \$1,410)	\$10,085 (\$0; \$11,227)	\$2,247 (\$0; \$2,410)	\$1,500 (\$0; \$1,500)
Facilities				
Site rental days	\$183 (\$100; \$300)	\$2,475 (\$1,350; \$4,050)	\$366 (\$200; \$600)	NA
Food and drink	\$272 (\$144; \$396)	\$3,575 (\$2,159; \$5,280)	\$544 (\$288; \$792)	NA
Subtotal	\$455 (\$244; \$696)	\$6,050 (\$3,509; \$9,330)	\$910 (\$488; \$1,392)	NA

Exhibit 10-3. Costs per deliberative session: Base estimate, low and high range estimates by method (2012 US dollars) (continued)

Cost Component	BCD: n = 12; 1 Day Base (Low; High)	CP: n = 24; 3 Days Base (Low; High)	CD: n = 12; 2 Days Base (Low; High)	ODP: n = 12; No Onsite Meeting Base (Low; High)
Equipment, Technology				
Microphones, speakers	NA	\$410 (\$250; \$580)	NA	NA
Projectors, screens	\$388 (\$375; \$400)	\$2,713 (\$2,625; \$2,800)	\$388 (\$375; \$400)	NA
Computers	\$2,097 (\$1,650; \$2,550)	\$2,097 (\$1,650; \$2,550)	\$2,097 (\$1,650; \$2,550)	\$2,097 (\$1,650; \$2,550)
Hard-line Internet connection in hotel	NA	\$467 (\$467; \$467)	NA	NA
Office-based Internet access	NA	NA	\$600 (\$600; \$600)	\$600 (\$600; \$600)
Conference line	NA	NA	NA	\$375 (\$375; \$375)
WebEx	NA	NA	NA	\$312 (\$312; \$312)
Subtotal	\$2,485 (\$2,025; \$2,950)	\$5,685 (\$4,992; \$6,397)	\$3,472 (\$3,000; \$3,950)	\$3,384 (\$2,937; \$3,837)
Materials				
Pens	\$1 (\$1; \$1)	3 (\$3; \$3)	\$1 (\$1; \$1)	NA
Paper and copies	\$101 (\$17; \$101)	\$932 (\$160; \$932)	\$147 (\$25; \$147)	NA
Flipcharts	\$50 (\$50; \$50)	\$550 (\$550; \$550)	\$100 (\$100; \$100)	NA
Easels	\$25 (\$25; \$25)	\$125 (\$125; \$125)	\$25 (\$25; \$25)	NA
Markers	\$16 (\$16; \$16)	\$80 (\$80; \$80)	\$16 (\$16; \$16)	NA
Subtotal	\$193 (\$110; \$193)	\$1,690 (\$918; \$1,690)	\$289 (\$168; \$289)	NA
Total	\$4,473 (\$2,379; \$5,249)	\$23,511 (\$9,419; \$28,644)	\$6,918 (\$3,656; \$8,041)	\$4,884 (\$2,937; \$5,337)

Note: We estimated low and high rates for all participant and facility inputs, as well as selected equipment, technology, and material inputs. Low estimates for participant inputs were set at \$0 to represent no reimbursement or incentive provided to participants. High estimates reflect a hypothetical reimbursement of \$15 for childcare and transportation costs. Low and high estimates are based on the costs at certain locations for the facility, equipment, technology, and material inputs. BCD = Brief Citizens' Deliberation; CD = Community Deliberation; CP = Citizens' Panel; NA = not applicable; ODP = Online Deliberative Polling.

Discussion

Many considerations affect the cost of implementing a deliberative method. In addition to considering the outcomes of deliberation that sponsors hope to achieve, sponsors must also weigh the feasibility of implementing different methods.

We show that the least costly method to implement on a per-session basis is BCD, with ODP and CD requiring a similar level of cost. As expected, CP sessions are the most costly to implement. The costs appear to be closely related to the number of hours required for each method. However, the implementation costs a sponsor will incur will also depend on how many groups are convened. A sponsor is likely to hold more than one group for any of the methods, but, in general, more groups would be convened with BCD, ODP, or CD than with the larger and more intensive CP method.

A number of factors may cause per-session implementation costs to vary. Sponsors seeking to decrease implementation costs could change equipment and technology requirements for sessions. For example, we used laptop computers, projectors, and screens to show presentations during sessions, but smaller scale projects could simply present information on paper.

We note that the “high tech” methods of ODP and CD, which include an online component, did not have substantially higher equipment and technology costs than BCD and CP. Although these methods rely on the Internet, the cost of Internet would generally be incurred by individual participants, facilitators, and experts at their homes or workplaces.

This analysis does not consider scalability. Per-person implementation costs for some methods will decrease as numbers of participants or groups increase because of the types of resources required. For example, ODP’s costs for conference lines and WebEx would be stable for projects that include more participants or groups.

Finally, sponsors must consider the costs we did not present when they project the full costs associated with fielding deliberative sessions. Important among these are the costs of materials, which can be kept quite low or made much more sophisticated. A related educational input, bringing in experts to present their perspectives on issues and answer questions, will sometimes be donated and sometimes be a significant cost.

Similarly, facilitation costs will vary substantially depending on a sponsor’s choice. Some organizations will draw upon and train their regular staff to provide facilitation, as we did. Other organizations hire consultants, and the costs for trained facilitators can vary markedly. Facilitation costs will also vary by method. CP often involves breakout as well as plenary sessions, so more than one facilitator is required. In ODP, the facilitators have the least intrusive role among the methods we demonstrated, so require much less training.

We intend this summary of implementation costs to help groups consider and plan for the resources required when selecting a deliberative method.

Section III: Implications

Chapter 11. Discussion and Implications

The goals of the Community Forum Deliberative Methods Demonstration were twofold:

(1) elicit public input on a topic of interest to AHRQ—the use of evidence in decisionmaking, and (2) expand the evidence base on public deliberation. In this chapter, we discuss findings and highlight implications relevant to entities that use medical evidence in decisions, as well as those interested in using deliberative methods to inform their work.

Public Input on the Use of Evidence in Decisionmaking

Our analysis of the public's input into the overarching deliberative question highlighted several areas potentially useful for entities that generate or use evidence in health care decisions.

The public's willingness and ability to apply evidence. The public can engage in discussion about ways to use and apply medical evidence when (1) relevant balanced background information is presented, and (2) they have the opportunity to discuss openly with others. Our findings support the conclusion that the topic of medical evidence is well within the public's purview for meaningful and relevant discussion. The combination of educational materials and interactive debate establishes an environment conducive to individuals increasing their knowledge and developing their views on how these topics affect them personally and as citizens.

The public's willingness and ability to take a broader perspective. The public is capable of viewing health care issues beyond the viewpoint of an individual; under certain circumstances, they prioritize societal needs over personal ones. To facilitate adoption of a societal perspective, people need information about the implications of different decisions for other individuals and/or for communities as a whole. Since the public has had limited opportunity to consider health care issues outside of their individual experience, explicitly asking them to consider the “community at large” sets the stage for accessing those social values relevant to complex health care issues.

Evidence of harms. Given that people have particular concerns about the impact of harms and are willing to accept constraints on their autonomy to let society take steps to avoid harm, effectiveness studies should be at least as attentive to this domain as they are to evidence of benefit. Researchers should consider whether they are currently giving adequate priority to studies focused on identifying harms or describing subpopulations or clinical situations for which interventions are harmful, and to efforts to communicate these findings. Messages about important harms, and especially harms to the larger community, should be highlighted when results are disseminated.

Limitations of using research evidence. Many concerns researchers and policymakers have about the limitations of the use of research evidence in decisionmaking are shared by the public. Specifically, the public's concerns were that (1) research done on populations that do not include people like them may not apply to them; (2) evidence is unreliable because it changes over time with new research; (3) in many situations, no good evidence is available; and (4) the evidence generated supports a specific self-interest. Some of these concerns have implications about how evidence is used and communicated. For example, understanding that evidence supporting a treatment option is limited to certain populations can help patients weigh the value of that evidence in discussing a treatment approach with their physician. Other concerns, such as the lack of evidence on specific health issues or patient populations, raise questions about research

methodology and approaches: how do researchers assess whether these gaps in evidence need to be addressed, and if so, are new approaches needed to address those gaps?

Perceptions of evidence. To members of the public, the term evidence covers not only the findings of research studies, but clinical judgment, test results, trial and error, and common sense. AHRQ and other organizations have funded much work in the areas of translation and dissemination of evidence; this finding highlights the complexity and inherent challenges in these efforts. Further, among the public, current levels of knowledge about evidence and comparative effectiveness are low. For comparative effectiveness research to be better valued, there would need to be a multifaceted approach that builds broader capacity and receptivity to this research and health care information.

Making clinical evidence relevant. Supporting the lay public's use and application of clinical evidence requires more than translating the results of clinical studies into plain language for the public. It also requires that evidence be put in the context of other factors—such as convenience, immediacy of results, or trust in one's practitioner—when presented to support personal health decisions. This presentation should show how the evidence applies to an individual patient and link their decisions with the outcomes that individuals want for themselves. For example, our examination of evidence supporting the effectiveness of high-performing hospitals highlighted the difficulty of using only the clinical evidence around surgical outcomes; some participants wanted information concerning other outcomes, including “patient-centered” respectful care, support for families, or having access to a local hospital that meets the community needs.

Transparency about financial motivations. Our participants were highly skeptical about the motivations of insurers, employers, government, and researchers involved in health care. Regarding insurers and employers, concerns centered on whether decisions aimed to increase profit at the expense of quality. For government, members of the public questioned why government should be involved in individuals' health care decisions; however, upon learning that the government is the largest payer, some participants were more willing to accept government involvement, given its need to contain health care costs. For researchers, concerns centered on conflicts of interest, in that funding sources might influence the conduct of research. Taken together, these findings suggest the importance of transparency when it comes to disclosing financial interests in health care overall, and specifically in the generation and use of evidence of medical effectiveness.

Expanding the Evidence Base on Public Deliberation

The results of the randomized controlled trial highlighted several areas potentially useful for those using or developing deliberative methods.

Deliberation: A Feasible and Effective Approach

Deliberative methods offer a feasible and effective approach for public and private organizations and institutions to obtain informed public views on complex topics affecting broader constituencies. Taken in aggregate, all deliberative methods elicited core values and, when compared with a reading materials only control group, demonstrated increases in knowledge and shifts in attitudes that showed that participation had meaningful effects.

Deliberation had similar effects for diverse groups. Public deliberation provides a framework in which individuals from varied social and cultural circumstances can interact effectively.

Deliberation had similar effects on people, no matter their race, ethnicity, age, or educational attainment. Diverse groups of participants reported learning from the experience of others, which allowed them to think about issues in a manner they would not have done previously.

An initial question we confronted in designing our approach was whether to conduct heterogeneous or homogeneous groups based on race and ethnicity. We decided that effective deliberation required heterogeneous groups of participants, allowing participants to share their perspectives with each other. Thus, in our study design, we not only recruited participants from different backgrounds to be part of each method, but also made sure to include participants from underrepresented groups. Overall, we posit that this recruitment strategy contributed to the effectiveness of deliberation and the similarity of effects across participants of different race, ethnicity, age, or educational attainment.

Deliberation provides different input from other types of public consultation approaches.

Deliberation allowed participants to explore their views in more detail, to see how information and context could shift discussion and elicit additional points of view, and to apply this information to the specific questions under discussion. Participants' comments and our data suggest that they had a deeper understanding of the issues under discussion; had stronger appreciation of the perspectives of research funders, payers, and their fellow participants; and generated solutions consistent with their core values. In contrast, surveys or opinion polls elicit intuitive, top-of-mind responses—individuals' initial beliefs and reactions concerning an issue. They are not intended to educate respondents or challenge them to consider their views. Focus groups give people an opportunity to interact with each other; however, in focus groups, the emphasis is on capturing a broad range of views, rather than on asking participants to explain reasons for their positions or consider others' viewpoints. Further, focus group participants are not typically asked to take a community or societal perspective.

Because of these differences, deliberation offers an advantage when members of the public are called on to examine complex social issues that involve competing values. Deliberation can be considered a tool to “see around corners,” providing valuable insight into what a more informed public might say. When making decisions about research, priorities, or policy, these informed views arguably provide an essential ingredient about public perceptions, values, and perspectives.

The choice of when to use public deliberation ultimately depends on the nature of public input desired and how that input will be used. Users of public input should also consider how to balance the more spontaneous input from surveys or focus groups with the more informed and considered input from deliberative efforts.

How to keep the public engaged. Deliberation has a twofold outcome; not only does it inform the sponsor about the public's perspective on a social value issue, but also the public feels positively about the opportunity to provide input. Sponsors of deliberation may consider what other opportunities would be available to build on this mutually beneficial experience. For example, sponsors may identify future deliberative events or other pathways for continued involvement for those people whose interest in policy and program decisions was piqued by participating in a deliberative event.

Selecting a Deliberative Method

Our findings have implications for those selecting a deliberative method. Planners will likely want to consider which types of outcomes are most important, as well as the investment required to implement the deliberative method. Exhibit 11-1 summarizes the findings from our study across different outcomes and provides our overall assessment of each method.

Exhibit 11-1. Overall assessment of outcomes and investment, by method

Method	Citizens' Panel	Community Deliberation	Online Deliberative Polling	Brief Citizens' Deliberation
Outcomes				
Changes in knowledge about evidence and comparative effectiveness research	Increase compared with reading materials control.	No difference compared with reading materials control.	No difference compared with reading materials control.	Increase compared with reading materials control.
Shifts in attitude about the use of evidence	Shifted 2 of 3 attitude measures compared with reading materials control.	Shifted 2 of 3 attitude measures compared with reading materials control.	Shifted 1 of 3 attitude measures compared with reading materials control.	Shifted 1 of 3 attitude measures compared with reading materials control.
Ability to elicit main themes and values from the public	Main themes and values emerged. Strong ability to explore nuances of themes.	Main themes and values emerged. Some ability to explore nuances of themes.	Main themes and values emerged. Limited ability to explore nuances of themes.	Main themes and values emerged. Limited ability to explore nuances of themes.
Participant-rated impact of experience	Highest rating	High rating	Lowest, but still high rating	High rating
Participant rated value of experience	Highest rating (tied)	Highest rating (tied)	Lowest, but still high rating	High rating
Resources				
Typical number of participants	24-30 per group	12 per group	12 per group	12 per group
Typical number of groups convened	1-2 groups	5-6 groups	20-30 groups (or 200-350 participants)	10-12 groups
Space required	In person	In person	None	In person
Other resources	Internet or phone connection if using remote experts	Internet connection between sessions	Audio and Internet-based conference services	None
Internet access for participants	Not required	Required	Required	Not required
Facilitation	Active facilitator; 3 facilitators per group	Active facilitator; 1 facilitator per group	Passive facilitator; 1 facilitator per group	Active facilitator; 1 facilitator per group
Implementation costs ^a	\$23,500 per group	\$6,900 per group	\$4,900 per group	\$4,500 per group
Overall assessment	Effective. Highly appropriate for complex topics.	Effective. Appropriate for complex topics.	Effective. May be more appropriate for less complex topics.	Effective. May be more appropriate for less complex topics.

^aThe costs we report are limited to those directly associated with holding deliberative sessions, including participant costs such as incentives or reimbursement for childcare or transportation; facilities costs such as site rental, food, and drink; equipment technology, such as microphones, projectors, Internet connection, and telephone conference lines; and supplies such as pens, paper, flipcharts, easels, and markers. Costs do not account for economies of scale when implementing multiple groups.

A number of other criteria are likely to be important in deciding what methods are most suitable:

- What is the nature of the deliberative question? The more complex or controversial the question, the more important it is that methods allow sufficient time for a variety of interactive activities and the opportunity to process the complexity of the issues and look at the topic from multiple perspectives.
- Does the topic require expert witnesses? While all methods require educational materials, the addition of experts affects the time and resources needed, particularly if real-time (in-person, online, or via Skype) interaction with the expert is preferred.
- Is the “public” of interest local, state, or national? The more localized the public of interest is, the easier it is to use in-person methods; the broader the population geographically, the more advantageous the online methods.
- Is this deliberation a one-time event or intended to be ongoing, with sessions conducted periodically with an established, or rotating, group of participants? If the latter, the model needs to be designed around particular features that sustain it for the long term.

Because all methods were effective to some extent in eliciting core values, increasing knowledge, shifting attitudes, and having an impact on participants, our overall findings indicate that there are many valid ways to conduct public deliberation. Our experience suggests that when designing deliberative methods, following the core principles of deliberation—convening a diverse group of participants, providing a balanced information base, and encouraging reasoning and equal opportunity for speaking—is important. Beyond staying true to these principles, planners who are developing or modifying methods to suit their needs and preferences can weigh the types of tradeoffs we identify and use our results to inform their choices.

Continuing the Learning: Areas for Future Research

To build on the results of the Community Forum Deliberative Methods Demonstration, additional studies of deliberation would be valuable to continue learning about the effectiveness of deliberation and how features of the deliberative process contribute to outcomes. Other studies comparing deliberative methods may want to:

- **Understand how mode and style of facilitation contribute to outcomes.** Our remote method, ODP, differed from the in-person comparators in facilitation style as well as mode. Understanding which of these factors drove the major differences between ODP and other methods would provide useful information for those interested in using online methods.
- **Comparatively examine deliberative outcomes with self-selected community groups as well as representatively recruited groups.** The Community Forum recruited participants to reflect the demographic diversity of the communities from which the samples were drawn. But many conveners of deliberation groups may want to make the process open to anyone who can attend in a given community. Further research could explore whether the outcomes that occurred in this study also occur, in either similar or different amounts, in self-selected groups that are open to anyone.

- **Explore the elements of case studies that work best to elicit social values.** In our initial versions of the hospital quality case study, we found that the case study did not engender sufficient tension to elicit different viewpoints from participants, and ultimately, their social values. However, introducing more complexity made our qualitative and quantitative findings more difficult to interpret. Understanding how elements of different case studies work best would provide useful guidance for those conducting deliberative sessions.
- **Assess the impact of deliberation versus a control group with no reading materials.** Our study was designed to isolate the impact of deliberation above and beyond reading materials. Including a no reading material control group could assess the total impact of deliberation.

In addition, our findings suggest that further research in the following areas would be valuable for public deliberation:

Shifts in attitudes as a measure of effectiveness. For effective deliberation to occur, participants need to be open and willing to consider other points of view and, when in line with their values, change their perspective and opinions on a topic. Yet measuring and interpreting these shifts in perspectives can be challenging, as participants' views change in many different ways throughout deliberation. Perspectives may become stronger, alternate across different viewpoints, or remain the same even after considering a different viewpoint. Additional research and approaches may be needed to understand how to best assess these different shifts in perspective and their impact on the outcomes of deliberation.

The long-term impact on participants. Although public deliberation obtains input from the public to inform public policy and programmatic decisions, many proponents have also highlighted the impact of deliberation on participant engagement as an important outcome. Participants in the Community Forum Deliberative Methods Demonstration reported an impact of deliberation on their opinions at the end of their deliberative session and showed tangible changes in their knowledge and attitudes within 1 to 2 weeks following deliberation. However, we were not able to capture whether deliberation affected their behaviors, such as asking questions of their doctors, taking an active role in their health care, or engagement in broader policy issues. Additional research could explore the long-term impact of deliberation on participants' views and how shifts in attitudes may lead to behavior change.

Promoting effective deliberation through online methods. While in-person methods had more impact on outcomes than the online method in our study, the appeal of Internet-based interactive communications will continue to grow as this platform becomes ever more sophisticated and accessible to all populations. A study that explored the effectiveness of different approaches to online deliberation would be valuable for guiding developments in this area.

Exploring how deliberation works in multistakeholder situations. By design, Community Forum groups excluded those with an economic or professional interest in the deliberative question, such as health care professionals, who could potentially overinfluence or severely inhibit discussion with their point of view as an expert in health care. The ability of the lay public to participate in deliberation that includes professionals with expertise or professional interests may be critical for many situations, such as involving patients and families in research, in hospital governance, or in community health assessment. An important question concerns how

we can support members of the public most effectively in these multistakeholder situations, and especially in situations that require learning about technical content.

Assessing the impact of deliberation on the sponsor. In our literature review of deliberative methods, we found that, although participants in public deliberation wanted to know how their input would be used, only a few case studies reported on how the results of deliberative procedures affected specific decisions, laws, policies, or practices.¹ Assessing the impact on the sponsor's decisions was beyond the scope of this project, but future research could explore the effectiveness of public input in terms of impact on the sponsor, as well as impact on participants.

Conclusion

Many organizations—researchers, health care providers, and public- and private-sector purchasers—as well as multistakeholder efforts to improve community health have an interest in capturing the public voice on value-laden health issues. Further, multiple topics raised by participants over the course of the demonstration concerning the financing, structure, delivery, and oversight of health care services are important issues undergoing transformations in concept and design at the local, State, and national levels. The Community Forum Deliberative Methods Demonstration found that public deliberation is an effective, feasible, and useful method to capture public input on these topics.

References

1. Carman KL, Heeringa JW, Heil SKR, et al. The Use of Public Deliberation in Eliciting Public Input: Findings From a Literature Review. (Prepared by the American Institutes for Research under Contract No. 290-2010-00005.) AHRQ Publication No. 13-EHC070-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2013. www.effectivehealthcare.ahrq.gov.
2. Burkhalter S, Gastil J, Kelshaw T. A conceptual definition and theoretical model of public deliberation in small face-to-face groups. *Comm Theory*. 2002;12(4):398-422.
3. Jacobs LR, Cook FL, delli Carpini MX. Talking Together: Public Deliberation and Political Participation. Chicago: University of Chicago Press; 2009.
4. Lukensmeyer CJ, Brigham S. Taking democracy to scale: creating a town hall meeting for the twenty-first century. *National Civic Review*. 2002;91(4):351-66.
5. Mansbridge J. Deliberative polling as the gold standard. *Good Society J*. 2010;19(1):55-62.
6. Mendelberg T. The deliberative citizen: theory and evidence. In: Delli Carpini MX, Huddy L, Shapiro RY, eds. *Research in Micropolitics: Political Decision Making, Deliberation and Participation*, vol 6. Amsterdam: JAI Press; 2002:151-93.
7. Gutmann A, Thompson D. *Why Deliberative Democracy?* Princeton: Princeton University Press; 2004.
8. De Stefano L. Facing the water framework directive challenges: a baseline of stakeholder participation in the European Union. *J Environ Manage*. 2010 Jun;91(6):1332-40. PMID: 20223583.
9. Carson L, Hartz-Karp J. Adapting and combining deliberative designs: juries, polls, and forums. In: Gastil J, Levine P, eds. *The Deliberative Democracy Handbook: Strategies for Effective Civic Engagement in the Twenty-First Century*. San Francisco: Jossey-Bass; 2005:120-38.
10. Miles MB, Huberman AM. *Qualitative Data Analysis*. 2nd ed. Thousand Oaks, CA: Sage Publications; 1994.
11. Tashakkori A, Teddlie CB, eds. *SAGE Handbook of Mixed Methods in Social and Behavioral Research*. 2nd ed. Thousand Oaks, CA: Sage Publications; 2010.
12. Webler T. "Right" discourse in citizen participation: an evaluative yardstick. In: Renn O, Webler T, Wiedemann P, eds. *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse*. Dordrecht: Kluwer Academic Publishers; 1995:35-86.
13. Sunstein CR. *The Law of Group Polarization*. John M. Oline Law and Economics Working Paper No. 91. University of Chicago Law School; 1999.
14. Farrar C, Fishkin JS, Green DP, et al. Disaggregating deliberation's effects: an experiment within a deliberative poll. *Br J Poli Sci*. 2010;40(2):333-47.
15. Kim SY, Uhlmann RA, Appelbaum PS, et al. Deliberative assessment of surrogate consent in dementia research. *Alzheimers Dement*. 2010 Jul;6(4):342-50. PMID: 20188635.
16. Min S-J. Online vs. face-to-face deliberation: effects on civic engagement. *J Computer-Mediated Comm*. 2007;12(4):1369-87.
17. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New Jersey: Lawrence Erlbaum; 1988.
18. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994.
19. Yankelovich D. *Coming to Public Judgment: Making Democracy Work in a Complex World*. Syracuse, NY: Syracuse University Press; 1994.
20. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 5th ed. Oxford, UK: Oxford University Press; 2001.
21. Mallery C, Ganachari D, Fernandez J, et al. *Innovative Methods in Stakeholder Engagement: An Environmental Scan*. (Prepared by the American Institutes for Research under Contract No. 290-2010-0005-C.) AHRQ Publication No. 12-EHC097-EF. Rockville, MD: Agency for Healthcare Research and Quality; May 2012. www.effectivehealthcare.ahrq.gov.

22. National Consensus Panel on Emergency Preparedness and Cultural Diversity. Guidance for Integrating Culturally Diverse Communities Into Planning and Responding to Emergencies: A Toolkit. Washington, DC; 2011.
23. Office of Minority Health. CLAS and the CLAS Standards. U.S. Department of Health and Human Services. www.thinkculturalhealth.hhs.gov/Content/clas.asp#clas_standards. Accessed October 23, 2013.
24. Bruni RA, Laupacis A, Martin DK, et al. Public engagement in setting priorities in health care. *CMAJ*. 2008 Jul 1;179(1):15-8. PMID: 18591516.
25. Sanders L. Against deliberation. *Polit Theory*. 1997;25(3):347-73.
26. Morgan DL. Focus groups. *Ann Rev Sociol*. 1996;22:129-52.
27. Boulware LE, Cooper LA, Ratner LE, et al. Race and trust in the health care system. *Public Health Rep*. 2003 Jul-Aug;118(4):358-65. PMID: 12815085.
28. Sofaer S, Firminger K. Patient perceptions of the quality of health services. *Annu Rev Public Health*. 2005;26:513-59. PMID: 15760300.
29. Nelson A. Unequal treatment: confronting racial and ethnic disparities in health care. *J Natl Med Assoc*. 2002 Aug;94(8):666-8. PMID: 12152921.
30. Fishkin JS. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford, UK: Oxford University Press; 2009.
31. Sunstein CR. The law of group polarization. *J Polit Phil*. 2002;10(2):175-95.

Appendix A. Technical Expert Panel

The Community Forum Technical Expert Panel comprised the following six experts in public deliberation, comparative effectiveness research, or risk communication. These experts provided guidance on implications for the design and implementation of public deliberation on comparative effectiveness topics.

Julia Abelson, Ph.D., Professor, Clinical Epidemiology and Biostatistics; Associate Member, Department of Political Science; Member, Centre for Health Economics and Policy Analysis (CHEPA), McMaster University

Stirling Bryan, Ph.D., Professor, Director, Centre for Clinical Epidemiology and Evaluation, University of British Columbia

Tim Carey, M.D., M.P.H., Director, Sheps Center for Health Service Research; Professor, Social Medicine and Medicine, University of North Carolina–Chapel Hill

Scott Y. Kim, M.D., Ph.D., Associate Professor of Psychiatry; Co-Director of Center for Bioethics and Social Sciences in Medicine, University of Michigan

Peter Muhlberger, Ph.D., Director, Center of Communications Research, Texas Tech University

David Ropeik, M.A., Independent Consultant and Instructor, Harvard University

Mark E. Warren, Ph.D., Professor, Harold and Dorrie Merilees Chair in the Study of Democracy, Department of Political Science, University of British Columbia

Appendix B. *Preparing for the Community Forum* Booklet

Preparing for the Community Forum:

Thinking about quality health care



Thank you for agreeing to take part in the Community Forum project!

This handout tells you about the Community Forum and what you can expect to do during this exciting project. This handout also gives some background information on quality health care and medical research.

What is the Community Forum?

The Community Forum is your opportunity to be part of a group that tells government decision-makers your ideas about how medical research can be used to improve the quality of health care for everyone. Your group's ideas will help federal government agencies make better decisions about how to improve health care.

Who is sponsoring the Community Forum?

The Community Forum is a project sponsored by a federal government agency called the Agency for Healthcare Research and Quality, or AHRQ. AHRQ works to improve the quality and safety of health care in the United States.

The American Institutes for Research (also called A-I-R) is a non-profit research organization that is leading the groups on behalf of AHRQ.

Why do we want to hear from you?

AHRQ often hears from doctors and researchers about health care. Even so, AHRQ feels that there is an opinion that is just as important as those of health care professionals—**Yours!**





What will happen during the group?

You will learn about and discuss how medical research can be used to make health care safer and better. The group takes place over three days. This may sound like a long time, but there will be breaks and lots of chances to talk and learn new information. A facilitator will lead the group, making sure everyone gets a chance to talk. You will get the chance to ask questions of experts such as doctors and researchers too.

What do I need to do?

- Come to the discussion with an open mind and be ready to listen, learn, ask questions, and share your ideas.
- AHRQ wants to know what you think is best for everyone in your community or society overall. It may be natural to think about yourself and your family first—that's important—but we also want to hear what you think is best when you think about everyone.
- We ask that you give reasons for your opinions or feelings, so everyone in the group can understand why you feel the way you do. Remember – your input is really important!
- Some of the issues we discuss may be new to you. Don't worry! We will help you understand the health topics, so you can be part of the discussion. There will also be plenty of time to ask questions.

Who is taking part in the group?

The group will include all types of people from your community to allow for a rich discussion and many different points of view.

What happens after the group?

We will put together the ideas we get from all of the groups across the country and write a report for AHRQ and other people who make decisions about health care. They will use everyone's input to make decisions about the use of medical research to improve the quality of health care.

Quality Health Care

What is good quality health care?

Good quality health care is care that gets people the best possible results for their health and well-being. Getting good quality health care can help people stay healthy and recover faster when they get sick.

According to the Institute of Medicine¹, good quality health care is:

- **Safe.** Safe health care is health care without medical errors. Doctors, nurses, and hospitals work hard to prevent medical errors, but mistakes still happen. A report by the Institute of Medicine estimates that as many as 98,000 people in the United States die in hospitals each year as the result of medical errors.
- **Effective.** Effective health care means patients get health care based on the latest evidence from medical research about what health care works best.
- **Patient-centered.** Patient-centered health care is when doctors, nurses, and other health care professionals respect and respond to the preferences, needs, and values of patients and their families.
- **Timely.** Timely health care is when patients get the health care they need at a time when it will do them the most good.
- **Efficient.** Efficient health care is health care that does not waste the patient's time or money—or cost more than it needs to.
- **Equitable.** Equitable health care is when everyone has the same chance to get good quality health care. It means people are not treated differently because of their gender, ethnicity, where they live, or how much money they make.



¹ The Institute of Medicine is an independent, non-profit research organization that is part of the National Academies of Sciences.



□ This figure shows the different parts of good quality health care, as described by the Institute of Medicine.

Medical Research and Medical Evidence

The Community Forum will get your ideas about how **medical research** and **medical evidence** can be used to make sure everyone gets good quality health care.

What is medical research?

Medical research is when doctors and researchers study groups of people to find out what types of health care work best for most people.

Everyone is different. But if medical research is done well, then doctors and researchers can feel sure that they have enough information to say if a treatment will work for most people.

What is medical evidence?

Medical evidence is when doctors and researchers have enough information from the results of medical research to say how well a treatment will work for most people.

Using medical evidence is part of good quality health care

When medical research is done well, it leads to medical evidence.

When there is medical evidence to show which health care works best, then using medical evidence is part of good quality health care.





Does everyone get health care based on medical evidence?

No. It may surprise you to learn that often Americans do NOT get health care based on the latest medical evidence.

Here are just two examples of research that have shown problems with quality of health care:

- Americans receive appropriate health care when they need it only 55% of the time. All Americans are at risk of receiving poor health care—no matter where they live, how much money they have, or their race, education, or health insurance.
- Over 90,000 Americans with conditions such as high blood pressure, diabetes, and heart disease die each year because they don't receive the type of health care that research has shown to work best for their condition².

What does it mean when medical research is done well?

In general, medical research is done well when:

- **The study has enough people** to say what the chances are of the same thing happening to other people.
- **The study lasts for a long enough time** to make sure all the benefits and risks of a treatment can be known. For example, one study may last for 6 months, but another study may last for 5 years.
- **Doctors and researchers design the study to reduce bias** so that they know the benefits and risks are due to the treatment and not to other factors.
- **More than one study shows the same results.** If more than one study shows the same results, then researchers can feel sure that the same thing will happen to other people.

² For details, see *The Essential Guide to Health Care Quality and State of Health Care Quality* by the National Committee for Quality Assurance (<http://www.ncqa.org/tabid/203/Default.aspx>).

Comparative Effectiveness Research

Medical research helps make sure people get good quality health care. There are different types of medical research. For this project, we will learn and talk about one type of medical research called **comparative effectiveness research**.

What is comparative effectiveness research?

- **Comparative effectiveness research**, sometimes called CER (C-E-R), is a type of medical research that compares treatments for the same health problem to see which one works best.
- **For example, researchers might compare two different asthma medicines to see which one helps patients breathe more easily.** Or, researchers might compare whether surgery or radiation works best to treat cancer.
- **CER does not help discover new treatments or services because it looks at treatments that are already available.** Researchers do other types of medical research to find new treatments.
- **Sometimes, there is not enough medical evidence from CER to say which kind of health care works best.** More research may be needed to find the answers.

Who does comparative effectiveness research?

Doctors and other medical experts do CER. These researchers work at different independent organizations, such as universities and public or nonprofit research centers. Government agencies, such as the Agency for Healthcare Research and Quality (AHRQ), pay to have much of this research done.

A Note About the FDA

The FDA (Food and Drug Administration) decides if a new drug or medical device can go on the market. The FDA makes sure that the drug or device does what it is supposed to do and is safe. But the FDA does not compare drugs or devices to find out which one works best for a specific health condition.

That's why it is important to look at the results of CER.

How can medical evidence from comparative effectiveness research be used?

Medical evidence from CER can be used in different ways. For example:

- **Doctors and patients can use the results of CER to choose the treatment that works best for the patient.** CER helps doctors and patients understand the facts about different treatments, such as what works well and what the problems may be.
- **Groups of doctors, who are experts in their field, may make quality recommendations for good patient health care.** These recommendations help doctors and other health care professionals know the best way to care for patients in most situations. Quality recommendations may also be called clinical practice guidelines or quality standards.

Health care that follows quality recommendations is sometimes called “evidence-based care” because it is based on medical evidence that says what type of health care works well.

- **Health insurance plans may use the results of CER to help decide what the cost of the treatment will be for patients.** For example, when medical evidence shows that one treatment works better than another, health insurance plans may set a lower cost to patients for that treatment to encourage patients to use it.



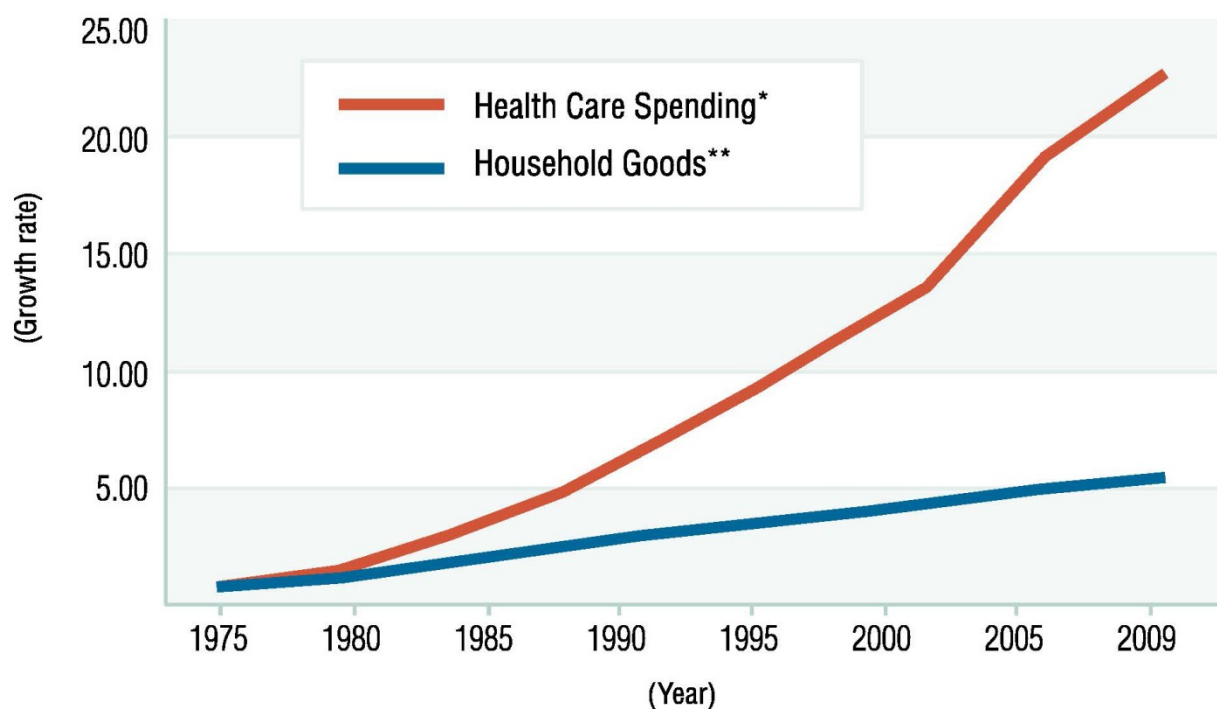
Health Care Costs

The costs of health care to society are important when thinking about the use of medical evidence to improve the quality of health care. After all, the wrong kind of health care can not only harm people, we also often spend more money than we need to.

Are health care costs increasing a lot?

Health care costs are increasing much faster than other parts of the economy. In 2009, health care spending in the United States totaled \$2.5 trillion. Health care spending in 2009 was 22 times more than in 1970. In comparison, prices for household goods were only 5.5 times more than in 1970.

Health Care and Household Goods Spending Rates, 1975–2009



* Health care spending data is determined by the National Health Expenditures per capita.

** Household goods data is determined by the Consumer Price Index. Sources: Centers for Medicare and Medicaid Services (CMS), Office of the Actuary; Bureau of Labor Statistics (CPI-U, U.S. city average, annual figures).

Who pays for our health care?

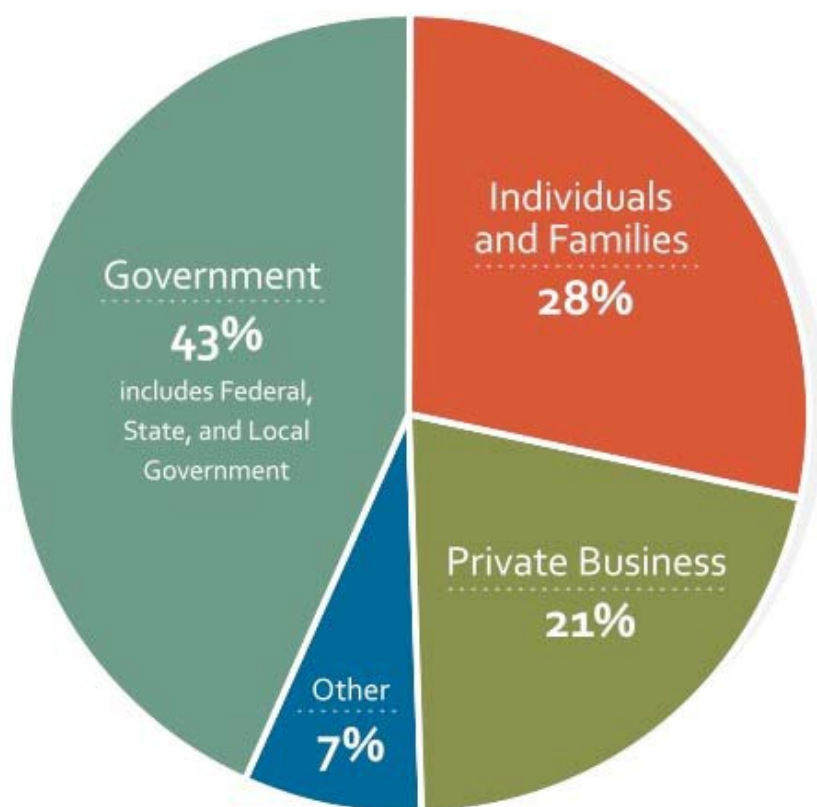
All of us play a role in paying for health care.

Individuals and families—even those who have health insurance—paid 28% of the costs of all health care in 2009. Individuals and families pay this amount in different ways. For example, they pay:

- Part or all of their insurance premiums
- Out-of-pocket costs for their medical care
- The portion of their salary that goes to support Medicare for seniors.

Also, the federal, state, and local government paid 43% of all health care costs in 2009.

Contributors To Health Care Payments in 2009*



* Estimates of spending by contributor are organized according to the underlying entity (business, households, and government) financing the health care bill payer. CMS refers to these contributors as “sponsors.” Figure does not add to 100 percent due to rounding.

** Other includes philanthropic giving, worksite healthcare, and revenues received by some health care providers for non-health activities.

Sources: Centers for Medicare and Medicaid Services (CMS), Office of the Actuary



For more information

If you would like more information about the group or any part of this project, please contact [contact information].

Appendix C. Case Studies

Comparing Hospital Quality

What is the health care issue?

Many serious health problems require treatment in the hospital. This treatment can be surgery or intensive medical care. As medical technology improves, hospitals can save more lives and treat more serious health problems. But doing more complicated medical work puts more pressure on hospitals to have highly skilled doctors and support teams, and increasingly complex equipment and facilities.

Hospitals that do a specific surgery or treatment for many patients are called high-volume hospitals for those treatments. Hospitals that provide that treatment for fewer people are called low-volume hospitals. A hospital that is high volume for one health treatment (for example, heart surgery) may not be high volume for a different one (for example, cancer surgery).

What does the medical evidence show?

Over the years, researchers have studied the differences in clinical results between high-volume and low-volume hospitals for a variety of health problems. Clinical results are things like complications from surgery, survival rates, and recovery times.

Because the teams of professionals providing care at high-volume hospitals have more experience, patients at high-volume hospitals sometimes get better clinical results than patients at low-volume hospitals. For example, in hospitals that do more surgeries, fewer patients get infections.

But, clinical results are not the only measure of a good hospital. Many people consider other parts of hospital care just as important in helping patients recover. These include how well the hospital staff communicates with and are respectful of patients and families, how well hospital staff manage a patient's pain, the comfort in receiving care from doctors and staff that they know, and how convenient the hospital location is for the patient and for family and friends to visit.

High-volume and low-volume hospitals: differences in clinical results

Children's heart surgery: In low-volume hospitals, 15% of the children on average will die from the surgery. In high-volume hospitals, 4% will die. This means that of 100 children having this surgery, 11 more would die in a low-volume hospital than in a high-volume hospital.

Artificial knee surgery: Knee surgery is not always fully successful. In a low-volume hospital, 5 percent of patients (5 in 100) will need to have new knee surgery within 6 years; 2 in 1,000 patients will die from the surgery. In a high-volume hospital, just 3 percent of patients will require new surgery, and 1 in 1,000 will die.

Surgery for aortic aneurysms (when the major blood vessel from the heart needs to be repaired): In low-volume hospitals, almost 7 in every 100 patients will die from this surgery. In high-volume hospitals, only 3 in 100 patients will die. Compared to other major surgery, this surgery is uncommon.

In treating more common health problems like **pneumonia, maternity care, and appendicitis**, low-volume hospitals are just as skilled as high-volume hospitals.

Comparing Hospital Quality



What do you think?

Given the differences in results for some patients in these low- and high-volume hospitals, which of the three statements is closest to your view?

- How hospital staff treat a person (such as being respectful, communicating well, etc.) is more important than the differences in clinical results.
- Regardless of what the evidence shows, people should trust that their doctors will provide the best care no matter which hospital they go to.
- People should only use the hospitals that get the best clinical results.

Comparing Hospital Quality

A closer look

Researchers and health care leaders are primarily concerned about providing the highest quality care to patients regardless of which hospitals they go to. But they also look at cost differences. In other words, when patients get better clinical results, does this cost more, less, or the same as for patients who do not?

Comparing the cost

Some studies compare the costs of patients who had complications from treatment to patients who did not have complications. For example, one research study compared patients who had major surgery and whether or not they had complications such as an infection, excessive bleeding, or blood clots that would slow down the patient's recovery. With this comparison, the cost difference was considerable:

- Patients without complications cost an average of \$28,000.
- Patients with complications cost an average of \$159,000.

These results suggest that if patients have more medical problems from their surgery, the cost of dealing with these problems can be significant.

Other studies compare the actual cost difference between low-volume and high-volume hospitals for particular surgeries. For example, a complex abdominal surgery costs \$27,000 in a high-volume hospital and \$33,000 in a low-volume. These cost differences may be because patients recover faster with fewer problems at a hospital that performs many of these operations over time.

Yet high-volume hospitals are not always less expensive. For example, patients with heart failure had better results at high-volume hospitals, but the cost of the care was greater than at low-volume hospitals.

Although patients who have health insurance may be protected from much of the extra costs that can come with less skilled care, these costs can have an impact on other people. That is because when more dollars are spent on health care, health insurance rates can increase for everyone.

Comparing Hospital Quality

The community of Springview

Imagine there is a small community, Springview, with one small hospital called Springview Community Hospital. This hospital is a low-volume hospital for most of its services, but the overall quality of its care has been pretty good.

A recent report showed that the Regional Medical Center, a hospital 50 miles away from Springview, had better clinical results for several types of surgery. The Regional Medical Center is a high-volume hospital for those types of surgeries.

The local government is the largest employer in the area. Half the residents get their health insurance through this employer. In the poor economy, the town can now only offer employees one health plan.

The town council has to decide which of two health plans it will offer:

Health Plan A (Low volume)

This plan requires that all patients use **Springview Community Hospital**, the local hospital.

Plan A only pays for care at the Regional Medical Center when patients need treatment that is not offered at Springview Community Hospital.

If employees wanted to go to the Regional Medical Center for a surgery where the Center gets better clinical results, it would not be covered by insurance.

Health Plan B (High volume)

This plan requires that all patients use the **Regional Medical Center**, 50 miles away.

Plan B only pays for services at Springview Community Hospital for two situations: 1) maternity care and 2) emergency care before transferring patients to Regional Medical Center.

If employees wanted to go Springview Community Hospital for a common health problem like pneumonia, it would not be covered by insurance.

Local hospital concerns

Springview Community Hospital was worried. If the county chose Health Plan B, it would lose many of its patients to Regional Medical Center. Without the income from local patients, some hospital services might have to be cut back. This could lead to layoffs for local residents. The hospital's board of directors started worrying about the hospital's long-term survival if Plan A is not chosen.

Local employees' reaction to these choices

Some employees and their families support keeping their medical care at the local, low-volume Springview Community Hospital. They want their family and friends close to give them emotional support when they are sick, avoid traveling long distances for hospital care, and maintain the comfort of having local doctors, nurses, and other staff whom they know and trust. They also didn't want to see the possible cut-back of local jobs.

Other employees support getting medical care at the high-volume Regional Medical Center. They did not want to jeopardize their health and that of their loved ones by agreeing to get medical services where the clinical results were not as good. They felt that having a small chance of a better outcome was more important than convenience, loyalty, or the possible threat of lost jobs.

What do you think?

Suppose all of you are on the Springview town council, which has to make the decisions for all the county employees and their families.

Which health plan would you pick and why?

As you read the case study, think about your answers to these questions:

- What is the right of the patient to make his/her own decision? If there is evidence that shows that decision (such as going to a hospital that gets poorer results) may not be in the best interest of the patient, is this still the patient's choice as long as the patient understands the risk?
- What is the authority of the doctor to make the treatment decision he/she thinks is best for the patient? If both the doctor and the patient agree to take a risk with a poorer-performing hospital, should that decision belong to them alone?
- What is your reaction to considering cost and hospital choice?
- Cost differences could have an impact on local employers who are paying most of the cost of the health insurance. Do those who bear the cost of purchasing the health insurance have any authority in deciding which healthcare facilities can be used?
- The town council decided to study the data closely in terms of how patients were affected who were treated at Springview Hospital. They learned that 6 patients last year (out of 500 total) had some complications that probably wouldn't have happened in a higher volume hospital and that one of these patients died from the complications.

Do these specific numbers affect your viewpoint? Would a higher or lower number of complications make a difference?

- Remember, the hospital's CEO and Board were worried that a significant loss of patients might harm the hospital's financial well-being. Imagine that Plan B was picked. Two years later, the town council had to again decide on the next health plan contract. By this time, the hospital was seriously in debt; dozens of local residents had lost their jobs; and the hospital board was now considering closing and selling the facility to a national nursing home chain.

Does this information affect your viewpoint?



Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment

What is the health problem?

Children get colds with symptoms like coughs and sore throats throughout their childhood. These conditions are called upper respiratory infections, or URIs. URIs are almost always caused by a virus. URIs caused by viruses will get better on their own within a week or so. But about 10 percent of URIs are caused by bacteria. Bacteria can cause “strep throat,” and doctors can do a quick test to see if this serious infection is present.

What to do about URIs?

Children with URIs have lots of uncomfortable symptoms. There are two approaches to treating URIs.

- **Relieve symptoms.** One approach to treat URIs is to relieve symptoms. This includes having a child drink plenty of fluids and, when necessary, use common drugstore medicines that the doctor recommends to loosen congestion, decrease cough, and control any fever.
- **Relieve symptoms and use antibiotics.** A second approach to treat URIs is to prescribe antibiotics. Although antibiotics have no effect on a URI that is caused by a virus, antibiotics work against a URI caused by bacteria. URIs caused by bacteria may lead to more serious illness if left untreated.

What does the medical evidence show?

When antibiotics are given to a child routinely, even for a good reason, bacteria can become “resistant.” This means the antibiotics are less likely to get rid of bacterial infections. Researchers believe that if a child is given antibiotics when not needed, those antibiotics might not work to stop future infections that the child may develop. This can put the child's health in danger when they really need an effective antibiotic.

Due to these concerns, the American Academy of Pediatrics established clinical practice guidelines more than 10 years ago. These guidelines say that to avoid overusing antibiotics, they should not be prescribed unless the doctor is sure that it is a bacterial infection and not a virus. The academy has taken steps to educate all doctors about these guidelines and the reasons for them.

What is actually done?

Most doctors follow the guidelines from the academy. But some doctors still prescribe antibiotics for URIs— even when they do not know if the URI is caused by a bacteria or virus. Doctors give various reasons for doing this:

- Although they can test for the strep bacteria, some doctors are concerned that there might be another type of bacteria causing the URI.
- Parents ask for antibiotics because they believe antibiotics help their child feel better sooner. Some parents need to avoid missing work to stay home with their child.
- Some doctors and parents think that if the antibiotics don't cause immediate harm to the child, it does not hurt to try them.



Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment

What do you think?

Which statement best reflects your view?

- If the doctor thinks that an antibiotic will help the child, then that is more important than what the evidence says.
- It is the parents' decision about getting the antibiotic or not; they care the most for their children, and they should decide if they think the benefit of the antibiotic is worth the risk or not.
- Since we know that too many antibiotics might harm the child in the long run, doctors should not prescribe antibiotics unless they know for sure the child has a bacterial infection.





Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment

What is the impact on the community?

As mentioned earlier, if antibiotics are used when they are not needed, they eventually become less able to kill bacterial infections. This affects not only the individual patient who has taken antibiotics when not needed. It also affects the larger population.

One example of this is a severe infection called MRSA (“mersa”). MRSA developed because a common antibiotic, a type of penicillin, was widely used to treat MRSA, and now it no longer works against the MRSA infection. It takes stronger, more dangerous drugs to control the MRSA bacteria. In 2005, the number of people hospitalized with the MRSA infection was more than 278,000, and about 5,500 people died that year due to this infection.

Due to examples like MRSA, public health physicians and the American Academy of Pediatrics have been trying for many years to help doctors and patients understand the problems of antibiotics. If antibiotics are overused, many people—children and adults—who need an antibiotic in the future may be in danger of getting an infection that cannot be controlled. But many parents find it difficult to put aside current worries about a child’s health for a **possible** problem that could impact the population at large in the future.





Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment

What do you think?

Now that you've learned more about overusing antibiotics, which statement best describes your view?

- I think it is up to the doctor and parents to decide if the antibiotic should be used in each individual case. Parents should not feel responsible for what might or might not happen in the future for the community as a whole.
- I think that given the risks to many people because of overuse of antibiotics, doctors and patients should consider the best interests of the community over individual interests and keep the use of antibiotics down.
- To avoid this problem, there should be stricter rules for when a doctor can order an antibiotic for a patient. We cannot always depend on people to do "the right thing" voluntarily.



As you read the case study, think about your answers to these questions:

- If the doctor and the parent both thought an antibiotic was needed, should anyone be able to object to this?
- Can you imagine any situations where the doctor's opinion should no longer be the 'final word'?
- If the parent wants it, should the doctor prescribe the antibiotic even if the doctor didn't believe it was helpful?
- Is it the doctor's duty to make sure the parents feel supported in what they believe is important—or to do what is in the long-term best interest of the child?
- If this situation applied to adults rather than children, would your views be any different?
- Guidelines on antibiotic use have been in effect for more than 10 years but some doctors still do not follow the guidelines. If providing all kinds of memos and reminders do not work, should the American Academy of Pediatrics just give up and accept that some doctors will never follow the guidelines? Or should there be some other actions taken?
- Can you imagine any situation that would cause you to decide that it isn't up to the individual doctor and patient? Or is an outbreak of a serious (and difficult-to-treat) infection the price we pay for the freedom we have to make our own medical decisions?
- The statement about 'stricter rules' suggests there should be real limits to doctors and patients having the freedom to choose the treatment they think is best in their situation. Are you ready to impose 'stricter rules' on other treatments that they order or request? Does a certain threshold have to be reached to impose strict rules? How would you describe what that threshold is?

Obesity Management: Comparing Treatment and Prevention

What is the health problem?

Obesity is a widespread and growing problem throughout the country. More than one-third of Americans are obese, as defined by their “body mass index,” or BMI. BMI is a number that describes a person’s weight in relation to their height. An adult who has a BMI of 30 and above is considered obese.

Obesity often leads to other medical problems such as diabetes, heart disease, arthritis, stroke, depression, and other long-term problems.

Compared with people of normal weight, people who are severely obese, with a BMI of 35 and above, are:

- Almost 2 times as likely to have heart disease
- Almost 4 times as likely to have high blood pressure
- 6 times as likely to have diabetes

	normal weight	overweight	moderately obese	severely obese
BMI	18.5–24.9	25–29.9	30–34.9	35 and above



What are the current approaches for treating or managing obesity?

Two main ways to manage obesity are (1) diet and exercise programs and (2) surgery.

Diet and exercise

Many programs are available to help patients lose weight by changing their diet and exercise behaviors. Other programs, often more intensive, help patients learn how to make changes in diet or exercise while addressing the emotional aspects of being overweight.

Few studies show long-term effectiveness of behavior change programs, but longer term programs are more effective than shorter term programs. Several studies have shown that patients in more intensive programs lose about 18 to 20 pounds within 6 months, but they regain 1/3 of their weight lost a year after the program.

Behavior change programs are low risk to the individual's health and can work with other treatments.

These programs often are not covered by insurance.

Surgery

Banding surgery is a type of bariatric surgery in which a band is surgically placed around the stomach to restrict its size, so individuals eat much less food than before. Before considering banding surgery, almost all patients have tried unsuccessfully to lose weight through diet and exercise programs.

Banding surgery offers a better success rate for losing weight than behavioral change programs. Studies have shown that patients lose more than 60 pounds with surgery. Also, research has shown that the amount of prescription medicines drops after surgery. For example, one study showed that the number of prescription medicines dropped from 21 to 13 for severely obese patients. Patients also experience fewer medical problems, such as high blood pressure.

Research is limited about the long-term success of weight loss. Surgery can pose mild to severe health risks. In a study of 299 patients, after 3 years of having the band, 88 percent, or 263 people, had one or more side effects or complications such as nausea and vomiting, band slippage, or stomach blockage. In the study, 25 percent, or about 75 of the patients, had their band removed.

The FDA (Food and Drug Administration) approved banding surgery for people who are severely obese and just recently approved banding surgery for people who are moderately obese if they also have a related health problem such as diabetes. This new approval means that more people could be eligible for banding surgery covered by their insurance.

What are current approaches for preventing obesity?

There are different ways to prevent obesity in a community. They include:

- Changing community environments, such as increasing green space, walkability via sidewalks or walking trails, or access to nutritious foods
- Changing school environments, such as increasing access to fruits and vegetables and physical education programs
- Changing policies, such as regulating food content or having “sin” taxes on non-nutritious foods.

What does the evidence show about prevention?

Evidence shows that community prevention efforts are promising, but more research is needed to know how policy or community changes could reduce obesity over time. Research shows that programs that increase access to fruits and vegetables at school, such as instituting salad bars, increase fruit and vegetable intake among students. School-based physical education has been shown to increase levels of physical activity and improve physical fitness.

Some researchers also use “stop smoking” efforts as evidence of the effectiveness of community prevention. Over the past couple of decades, “stop smoking” efforts in the community have reduced smoking rates by half and saved resources.

Although more research is needed on the effectiveness of prevention, research shows that certain factors will increase the risk of obesity for people living in a community. These include:

- Not having markets near them that provide fresh fruits and vegetables
- The cost of fresh fruits and vegetables
- Local stores having processed foods with high sugar, salt, and fat content
- Too many fast food restaurants
- Not having “green space” or safe areas for leisure exercise
- The cost of using private gyms and exercise clubs

These risks for obesity may affect low-income, minority, and rural communities more than other groups. The United States Department of Agriculture estimates that 23.5 million Americans, half of whom are low income, live in food deserts, which are urban or rural areas without access to fresh, healthy, and affordable foods.

Which approach has priority?

You are a part of a community group that has just been given a federal grant to tackle the problems of obesity in your community. The grant allows you to do 3 of the following 6 options:

1. **Intensive diet and exercise programs** for people who are moderately obese (treatment)
2. **Surgery** for people who are moderately obese (treatment)
3. **Change the community environment** by increasing green space, walkability via sidewalks or walking trails, or access to nutritious foods (prevention)
4. **Change the school environment**, by increasing access to fruits and vegetables and physical education programs (prevention)
5. **Increase taxes on non-nutritious foods for people who buy them**, by changing policy to make it harder for people to buy these unhealthy foods (prevention)
6. **Increase taxes on companies who make non-nutritious foods**, by changing policy to make companies who make unhealthy foods pay for the costs of harms (prevention)

Decide which 3 options you think are most important for managing obesity in your community.

As you read the case study, think about your answers to these questions:

- What are your thoughts on how well diet and exercise programs work? Considering that patients in these programs do not lose as much weight as they do with surgery but experience lower health risks, should more or less funding go to such programs?
- What about banding for those who are moderately obese? Even if this is the most effective intervention for reducing weight and reducing the medical problems associated with obesity, do the benefits outweigh the risks?
- In thinking about ways to prevent obesity, what are your thoughts on the local community plan presented? What makes the local community plan be more or less appealing than the treatment programs suggested?
- What are your thoughts on the effectiveness of the local community programs in comparison to their feasibility?

Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicines

What is the health problem?

Coronary heart disease is the leading cause of death for all Americans. Heart disease occurs when the blood vessels in the heart become clogged, and blood and oxygen have a hard time getting to the heart. Not getting enough blood or oxygen to the heart can cause a range of problems, such as chest pain (angina), shortness of breath, limitation of activities, a heart attack, and death.

Heart disease ranges from mild to severe. People with mild heart disease have some blockage in their blood vessels, but the symptoms—mainly chest pain—are not severe. People with mild heart disease need to be treated to relieve chest pain and to prevent blockages from becoming worse.

What are the treatments for mild heart disease?

There are two main ways to treat mild heart disease:

- **Medicines only.** Doctors prescribe a combination of medicines, such as those that lower cholesterol, control blood pressure, relieve chest pain, reduce blood clotting, and others. Medicines relieve symptoms, such as chest pain or shortness of breath, but a small percent of patients will not have as good relief from medicines as from a stent. Medicines pose possible risks to the liver and kidney, which can be checked regularly with lab tests.
- **Stent plus medicines.** A stent is a metal device that is placed into a narrowed or clogged heart vessel to keep it open so that blood flow to parts of the heart is improved, or protected from further damage. Stents relieve symptoms, such as chest pain and shortness of breath, as soon as they are put in place. But, stents have more risks than only taking medicines. For example, 1 out of 100 patients will have a heart attack from getting the stent. And, 2 out of 1,000 patients will die directly related to having the stent put in. Patients who get a stent also take medicines like those described above.

All patients with mild heart disease are advised to make lifestyle changes, such as stopping smoking, limiting salt and fat in their diets, and exercising regularly, as long as the patient can do so safely.

What does the medical evidence show about each treatment?

Researchers compared the two types of treatment:

- The research showed no difference between the two treatments in the number of deaths or heart attacks over a period of 5 years.
- Getting a stent has more risks than medicines only, but stents bring faster relief from symptoms than medicines only. People who get a stent have fewer symptoms and report a higher quality of life than people who only take medicines. But, after 2 years, these differences go away.

Based on this research, only one out of three patients, or 33 percent, eventually will need a stent. The rest will have comparable results from medicines only.

What is the research recommendation?

Based on the medical evidence, medical experts recommend that medicines should be the **first** treatment given to patients with mild heart disease. If patients continue to have chest pain and other symptoms of mild heart disease, then experts recommend offering a stent.

What is actually being done?

A major study showed that 55 percent of patients were getting stents as the first treatment, rather than getting medicines only as recommended. This number is significantly more than the 33 percent of patients who need it.

This study raised concerns that many patients are getting stents that they do not need.

What are the costs of treatment?

With so many doctors doing stent procedures, researchers looked at the difference in what these treatments cost. Assuming all patients were also on medicines, then the only difference in cost would be the cost of performing the stent procedure in patients when there was no clear need for them.

In 2004, there were more than 650,000 stents placed as the first treatment in patients with mild heart disease. The average cost of placing a stent was \$56,000. This means that in 1 year (excluding those who would have needed a stent eventually), \$37 billion was spent on this procedure that research found to be “not necessary.”



Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicines

As a reminder, 33% of mild heart disease patients should get stents as the first line of treatment based on the medical evidence. But, 55% of patients are getting the stent first, meaning that doctors are doing almost twice as many stents than needed for mild heart disease.

What do you think?

You are a patient advisory group to a health plan (insurance company). The health plan wants to make sure that the number of patients getting stents is close to what the medical evidence says is needed for mild heart disease.

Your job is to advise the health plan on whether or not to reduce the number of stents being placed.

What would you recommend and why?



Comparing Approaches To Preventing Illness: a Fictional Case

Earth has recently received visitors from Asteroid 2WK7. These aliens, who are similar to humans, bring a new illness with them as they intermingle with earthlings. You are asked to come up with recommendations about how to manage this new illness. The information we have about preventive and treatment strategies are from comparative effectiveness studies from that Asteroid, where you can be reassured that the research is topnotch.

What is the health problem?

Shake, Rattle, and Roll, or SRR, is an illness that begins gradually with a mild trembling to all parts of the body, and people tend to break into quiet singing. Over a period of years, people will develop severe symptoms such that they can no longer conduct their normal activities due to the trembling and loud singing during all waking hours. SRR is not physically painful, does not affect a person's mental capacity, and does not cause death. But, as the disease progresses, patients become completely dependent on others for their personal care.

SRR will strike 20 of every 1,000 people. Everyone who gets SRR will develop the severe symptoms over a period of years. (As a reference, at any one time, diabetes affects about 80 of 1,000 people, and cancer affects 40 of 1,000 people.) The average cost of full-time care for people with SRR is \$100,000 per person.

What is the best way to manage SRR?

Experts have identified three approaches to manage SRR, but they disagree about which approach works best.

Option 1: Focus efforts on treating those who get SRR.

SRR can be treated with the medicine "Do Not Tremble." Half of the people who take this medicine will be cured, but the other half will eventually develop severe symptoms. "Do Not Tremble" is expensive to make because the ingredients have to be imported from Asteroid 2WK7. The total cost of treatment will be \$25,000 per person.

The focus on treatment will affect society in such a way that for every 1,000 people:

- A total of 980 people will not get SRR.
- A total of 20 people will get SRR and need treatment.
- 10 of those who get treatment will still have SRR and need full-time care.
- Cost to society for every 1,000 people is \$1.5 million.
(Costs include treatment for 20 people and full-time care for 10 people.)

Option 2: Focus on prevention by encouraging everyone to exercise 30 minutes per day.

Research has shown that people who exercise a minimum of 30 minutes per day will not get SRR. People who do not exercise still have a chance of getting SRR. Public health officials estimate that only 50 percent of people will exercise every day, even when they know the importance of doing so. Thus, exercise will reduce the number of people getting SRR by half.

Although exercise has no direct costs associated with it, public health officials estimate that it will cost \$1,500 per person to train everyone how to exercise correctly. Assuming that half of the population will do the needed exercise, the impact on society for every 1,000 people will be:

- A total of 990 will not get SRR.
- A total of 10 people will get SRR and need treatment.
- 5 of those who get treatment will develop SRR and need full-time care.
- Cost to society for every 1,000 people is \$2.25 million.
(Costs include exercise training, treatment for 10 people, and full-time care for 5 people.)

Option 3. Focus on prevention by putting medicine in the water supply.

"Prevent" is a preventive medicine for SRR that is placed in the water supply. No one who gets "Prevent" will come down with SRR. But, for every 1,000 people who get "Prevent," one person will have an allergic reaction to the medicine and eventually die from it. "Prevent" is estimated to cost \$5 per person when placed in the drinking water.

Using "Prevent" in the water supply will affect society in such way that for every 1,000 people:

- A total of 999 will not get SRR.
- One person will have an allergic reaction and eventually die.
- Cost to society for every 1,000 people is \$5,000.

What do you think?

Knowing these facts about SSR, which option would you choose?

_____ I would recommend **Option 1: Focus on treating those who get SRR.**

Expecting people to exercise every day is unrealistic. And, the risk of one person in 1,000 dying from putting medicine in the water supply is high.

_____ I would recommend **Option 2: Focus on prevention by encouraging everyone to exercise 30 minutes per day.** People should take responsibility for their own health. Fewer people will get sick without the risk of dying from an allergic reaction.

_____ I would recommend **Option 3: Focus on prevention by putting medicine in the water supply.** The benefit of no one coming down with a severe illness outweighs the small risk of one person in 1,000 dying from an allergic reaction. The financial costs to society are minimal.

Appendix D. Screenshots of the Deme Web Site

In this appendix, we present screenshots of the Deme Web site as an example of how we used this Web site during the Community Forum Deliberative Methods Demonstration.

Welcome: Home Page and Resources

This page briefly described the Community Deliberation group and provided links to resources, such as materials and instructions; tabs for information about the experts, a discussion board, and polling Web pages; and a resource box with the handouts and materials from the first in-person meeting.

The screenshot displays the 'COMMUNITY FORUM' website. At the top, a banner states 'This project is funded by the Agency for Healthcare Research and Quality (AHRQ)'. The header includes the AHRQ logo, the 'COMMUNITY FORUM' title with three stylized person icons, and navigation links: 'Welcome', 'Meet the Facilitators', 'Meet the Experts', 'Discussion Board', 'Take the Poll', and 'Help'. On the right, there are links for 'Decrease font size', 'Increase font size', and 'Click here to log out'.

The 'Resources' sidebar on the left lists several documents: 'Preparing for the Community Forum', 'Upper Respiratory Infections Case Study Part 1', 'Upper Respiratory Infections Case Study Part 2', 'Hospital Quality Case Study', 'Dr. Kleinman's Perspective', and 'Ms. Mitchell's Perspective'. Below this list is a printer icon and a link to 'Click here to get all documents on this page in one download'.

The main content area features a 'Welcome to Community Forum!' message, thanking users for joining the first session. It lists resources available on the site: 'Meet the Experts' (introducing Ms. Mitchell and Dr. Kleinman), 'Read Dr. Kleinman's perspective and Ms. Mitchell's perspective' on the issues, 'Discussion Board' for posting comments and questions, 'Take the Poll' about evidence and antibiotic use, and 'Read the case study on Hospital Quality' for the final session. A 'Help' link is also provided for navigation assistance.

Technical support contact information for Alex Ortiz is provided: communityforum@ahrq.org or 1-866-236-4285. A 'Thank you!' message follows.


On the right, a video player titled 'Watch the Welcome Video from AHRQ's Director, Dr. Carolyn Clancy' shows an 'Introduction by Dr. Carolyn Clancy'. Below the video, a link is provided: 'If you would like to see this video with Spanish subtitles, please click [here](#).'

At the bottom left, a section titled 'Do you need Adobe Software?' offers links to 'Get Adobe Reader' and 'Get Adobe Flash Player'.


Meet the Experts

This page showcased each expert's biosketch, photo, and brief statement on the deliberative issue(s) being discussed.

This project is funded by the Agency for Healthcare Research and Quality (AHRQ)



COMMUNITY
FORUM




Decrease font size | Increase font size

Click here to log out

Resources

- Preparing for the Community Forum
- Upper Respiratory Infections Case Study Parts 1 and 2
- Hospital Quality Case Study
- Dr. Kleinman's Perspective
- Dr. Sabin's Perspective
- Instructions for Using the Web site



Click here to get all documents on this page in one download

Do you need Adobe Software?


- Get Adobe Reader
- Get Adobe Flash Player

WelcomeMeet the Facilitators**Meet the Experts**Discussion BoardTake the PollHelp

Meet the Experts


Our two experts will be online to answer questions, respond to comments, and offer other perspectives on the topics. The experts will come online at least once a day to check out what people have been saying and add their own responses and comments.

Here are the bios and photos of the experts who will join our discussion on this Web site:



Dr. James E. Sabin

James E. Sabin, MD is a Clinical Professor in the Departments of Population Medicine and Psychiatry at Harvard Medical School and Director of the Harvard Pilgrim Health Care Ethics Program, a nationally recognized, highly innovative organizational ethics program. His major research interests are in the clinical and theoretical aspects of fair resource allocation, the role of consumers in overseeing health policy and practice, and the ways in which organizations can best deal with the ethical dimensions of their work. He blogs about health system ethics at <http://thehealthcareorganizationalethics.blogspot.com/>. To read about Dr. Sabin's thoughts on the antibiotic use and evidence, please click [here](#).




Dr. Lawrence Kleinman


Lawrence C. Kleinman, MD, MPH is Vice Chair for Research and Education of Health Evidence & Policy and Associate Professor of Health Evidence & Policy and Pediatrics at Mount Sinai School of Medicine, and Adjunct Associate Professor of Society, Health, and Human Development at Harvard School of Public Health. He is a leader in the field of children's health quality. To read about Dr. Kleinman's thoughts on antibiotic use and evidence please click [here](#)

Discussion Board

On this page, participants viewed ongoing discussions, replied to a comment, or started discussion topics (threads). Participants could post their discussion topics and responses at their leisure. All topics and responses were visible only to participants, facilitators, and observers within their Community Deliberation group.

This project is funded by the Agency for Healthcare Research and Quality (AHRQ)



COMMUNITY FORUM



[Welcome](#)
[Meet the Facilitators](#)
[Meet the Experts](#)
[Discussion Board](#)
[Take the Poll](#)
[Help](#)

[Decrease font size](#) | [Increase font size](#)

[Click here to log out](#)

Resources

- Preparing for the Community Forum
- Upper Respiratory Infections Case Study Part 1
- Upper Respiratory Infections Case Study Part 2
- Hospital Quality Case Study
- Dr. Kleinman's Perspective
- Ms. Mitchell's Perspective



[Click here to get all documents on this page in one download](#)

Do you need Adobe Software?

- Get Adobe Reader
- Get Adobe Flash Player

Community Forum Discussion Board
[Create a new discussion](#)


Discussion Topics	Replies	Last Post
Why is healthcare in the media? By margeg	0	N/A
Obamacare By tomh	0	N/A
Test By tomh	0	N/A
Who should make decisions? By Dr.Kleinman	0	N/A
help me understand By selbys	3	By davidd July 24, 2012, 1:15 p.m.
Leaking Calories By nydiam	0	N/A
How much cake can I eat and not gain weight? By kimbers	1	By kimbers July 23, 2012, 1:25 p.m.
Why if the Fed govt oming between me and my ... By jasonf	0	N/A
Do these health care decisions differ across families? By Dr.Kleinman	0	N/A


[Create a new discussion](#)

Take the Poll

This page listed three responses to the issue that participants discussed during the first in-person Community Deliberation meeting, which participants could vote on by agreeing or disagreeing. After submitting their responses, participants could post additional views on the statements by submitting their comments in a text box.

This project is funded by the Agency for Healthcare Research and Quality (AHRQ)




**COMMUNITY
FORUM**

[Welcome](#) [Meet the Facilitators](#) [Meet the Experts](#) [Discussion Board](#) [Take the Poll](#) [Help](#)

[Decrease font size](#) [Increase font size](#) [Click here to log out](#)

Resources

- [Preparing for the Community Forum](#)
- [Upper Respiratory Infections Case Study Part 1](#)
- [Upper Respiratory Infections Case Study Part 2](#)
- [Hospital Quality Case Study](#)
- [Dr. Kleinman's Perspective](#)
- [Ms. Mitchell's Perspective](#)



[Click here to get all documents on this page in one download](#)

Take the Poll

Now that you've learned more about overusing antibiotics, please read over the three statements below. Click the "Agree" or "Disagree" button beneath each statement.

1) I think it is up to the doctor and parents to decide if the antibiotic should be used in each individual case. Parents should not feel responsible for what might or might not happen in the future for the community as a whole.

☐ Agree ☐ Disagree

2) I think that given the risks to many people because of overuse of antibiotics, doctors and patients should consider the best interests of the community over individual interests and keep the use of antibiotics down.

☐ Agree ☐ Disagree

3) To avoid this problem, there should be stricter rules for when a doctor can order an antibiotic for a patient. We cannot always depend on people to do "the right thing" voluntarily.

☐ Agree ☐ Disagree

[Submit](#)

Do you need Adobe Software?

- [Get Adobe Reader](#)
- [Get Adobe Flash Player](#)

Appendix E. Final Qualitative Code List

CONTENT DIMENSION 1. Actor role within the health care system

Code #	Code name	Code description (inclusion and exclusion)
1.0	Actor role within the health care system	Avoid using this global code; instead tag comments with as many actor codes as are applicable. However, if a comment truly covers the entire gamut of actors listed below, you may place the comment in this code. This should be very rare.
1.1	Societal role and responsibility	Comments that address the role of society in health and health care. Comments may be about an actual role or about perceptions/opinions about what society should do. This code is for comments in which people talk about society generically, rather than a particular social institution such as government or the education system. You may also include the truly vague "They should" comments here if you really can't tell to whom the participant might be referring.
1.2	Individual/patient role and responsibility	Comments that address the role and responsibilities of the individual regarding health and health care. Comments may be about an actual role or about perceptions/opinions about what individuals should do. Include comments about personal accountability, such as comments about exercising/eating right/getting preventive care or comments about smoking/not exercising/eating badly/not getting preventive care.
1.3	Health care provider role and responsibility	Comments about the role or responsibilities of health care providers, such as comments about doctors having autonomy versus being answerable for socially imposed limits. For comments about professional accountability, consider whether it is appropriate to double-code those comments with "Boundary-setting mechanisms" (i.e., is the accountability to a professional organization as opposed to self-imposed or abstract?). Put comments about provider-patient relations, interactions, and communication under 3.7, "Doctor-patient relationship." This code is for providers as individuals. Put comments about institutions, such as hospitals or clinics, under 3.6, "Health care facilities or services." Double-code comments about doctor's ulterior motivations (taking money from pharmaceutical companies, making care decisions on a financial basis, etc.) under 5.2, Trust.
1.4	Government role and responsibility	Comments and discussion related to the roles and responsibilities of government (any level) in health, health care, and health policies. Include comments related to roles/responsibilities of institutions that are primarily government, such as schools.

CONTENT DIMENSION 1. Actor role within the health care system (continued)

Code #	Code name	Code description (inclusion and exclusion)
1.5	Other roles and responsibilities	Comments about the role of the legal system, professional organizations, etc. Use narrower codes 1.5a–1.5c whenever possible. Any other parties (other than care providers, individual patients, government, and society) mentioned as having a role in connection with health care or health care decisionmaking should be tagged with this code.
1.5a	Other—insurance	Comments mentioning insurance as a role player in the health care system or in health care decisionmaking.
1.5b	Other—pharmaceutical industry	Comments mentioning pharmaceutical companies as role players in the health care system or in health care decisionmaking.
1.5c	Other—family	Comments mentioning family members as having an important role in the health care system or in health care decisionmaking.
1.6	Society—definition	Comments that specifically address definitions of society, including comments about levels of society (e.g., my community versus the State or Nation). Include comments about the difficulty of defining society, such as determining where society ends and government begins. Include comments about society being changeable, abstract, or complex.

CONTENT DIMENSION 2. Boundary-setting mechanisms

Code #	Code name	Code description (inclusion and exclusion)
2.0	Boundary-setting mechanisms	Comments about boundaries and limits when a narrower subcode does not apply. Include information about regulating guidelines for medical practice. Include self-imposed boundaries and limits, and limits created by economic factors such as the ability to pay.
2.1	Penalties	Comments about penalties to discourage particular choices or practices. Penalties may be imposed by government, professional societies, institutions, insurance companies, etc.
2.2	Incentives	Comments about incentives to encourage particular choices or practices. Incentives may be provided by government, professional societies, institutions, insurance companies, etc.
2.3	Education	Comments about patient or provider education that will lead people to make the right choices.
2.4	Guidelines, regulations, rules	Comments about the use of, desire for, or objection to guidelines, rules, regulations, etc., that apply to health or health care. These guidelines may issue from any source (government, professional organization, insurance company, or the generic “they”). If a source of regulation/guideline is clear, please also code with the appropriate “Actor role” code (1.).

CONTENT DIMENSION 3. Decisions and decisionmaking

Code #	Code name	Code description (inclusion and exclusion)
3.0	Decisions and decisionmaking	Comments that do not fit into a narrower subcode.
3.1	Weighing the evidence	Comments in which participants discuss weighing the evidence as part of decisionmaking. This encompasses weighing evidence from competing sources, weighing research evidence against anecdote or personal feelings, and weighing evidence against values/ethics considerations. Use for comments related to effectiveness of treatments or policies and for comments about uncertainty of evidence. This code should capture participants' thoughts about the impact of evidence on the decisionmaking process. This is the use of evidence before a decision is made.
3.2	Weighing impacts and outcomes (short-term, long-term)	Use this code for comments that specifically discuss weighing the impacts or outcomes for individuals and/or society. These may be health-related (e.g., improvement in patient outcomes) or not (e.g., loss of access to facilities in a community). This code should capture participants' thoughts about how to weigh the impacts and outcomes of decisions that are made. This encompasses the consideration of impacts resulting from a decision that will be/has been made.
3.3	Cost	Comments about financial costs of all types, individual and societal. <ul style="list-style-type: none"> • Actual cost of treatments/procedures • Out-of- pocket costs, reimbursements • Insurance premiums, coverage • How to fund or pay for health systems (e.g., public health initiatives) • Medicare/Medicaid • Source of payment
3.4	Prevention	Use this code for comments that specifically relate to prevention as a goal. Double-code with 5.3, "Priorities," as needed.
3.5	Treatment	Use this code for comments that specifically relate to treatment of medical conditions as a goal. Double-code with 5.3, "Priorities," as needed.
3.6	Health care facilities or services	Use this code for comments that focus on medical facilities and services, such as hospitals, doctors' offices, emergency services.
3.7	Doctor-patient relationship	Comments discussing the doctor-patient relationship in connection with decisionmaking. Use for substantive comments discussing the nature of the relationship and its impact on decisionmaking. Do not code minor comments such as "I have a good relationship with my doctor" or "the relationship is important."

CONTENT DIMENSION 4. Values and ethics

Code #	Code name	Code description (inclusion and exclusion)
4.0	Stated values, ethical principles, beliefs, and preferences	<p>Comments in which participants state beliefs, values, ethical principles, or preferences about health, health care, health care use, or social or governmental involvement. Include explicit comments such as “I think education is the number one key” or “That information should only be given to X type of patients and not shared with Y patients.” Also include comments where the belief/value/principle/preference is obvious, such as “I assume that people who get this procedure are on public assistance,” or “There’s no restraint or accountability in the medical professions for ordering lots of tests,” or “We need independent evaluators of health care services.” For thematic categories that may appear, we may develop subcodes:</p> <ul style="list-style-type: none"> • More care is better care • Patient and/or provider education will solve the problem/lead people to the right choice • People have the right to choose badly • Guidelines are inflexible and take away choice • Convenience and rapid access are important (patient as customer) <p>Do not code comments in which you are not sure whether a value/principle/belief/preference is being expressed. If a value is not clear and obvious to you, do not code it.</p>
4.1	Choice, freedom	<p>Comments specifically about having or making choices, or about having the freedom to choose.</p> <p>Include comments about “good” or “right” choices and “bad” or “wrong” choices.</p>
4.2	Protection from harm	<p>Use this code for comments that specifically address the notion of protection from harm. Participants will often use this exact phrase, but near equivalents should also be tagged with this code.</p> <p>Double-code with 5.3, “Priorities,” as appropriate.</p>
4.3	Beliefs about medical condition, situation, or disease	<p>Comments indicating that participants hold particular beliefs about a condition, such as “It is avoidable” or “It is life threatening.”</p> <p>Include comments indicating that participants hold particular beliefs about patients with the condition, such as “Obese people are seeking quick fixes and seek out surgery before trying alternatives.” Also include comments about “everybody’s being different” or treatments working for some people, but not others.</p>
4.4	Respect, caring	<p>Use this code for comments that reflect an emphasis on the values of respect and caring and on the quality of life. This code is for comments that are not directly related to health and health care; those comments should be coded with the appropriate 1.0 “Role” codes and/or 3.0 “Decisionmaking” codes.</p>

CONTENT DIMENSION 4. Values and ethics (continued)

Code #	Code name	Code description (inclusion and exclusion)
4.5	Self-reliance, personal responsibility	<p>Use this code for comments that reflect an emphasis on autonomy, individualism, and personal responsibility. This code encompasses the philosophy of “pulling yourself up by your bootstraps” and not relying on other people or organizations. It also encompasses the belief that people need to “do what they are supposed to,” to be good citizens (or parents or spouses or employees, etc.).</p> <p>This code is for comments that are not directly related to health and health care; those comments should be coded with the appropriate 1.0 “Role” codes and/or 3.0 “Decisionmaking” codes. It may be difficult to separate these comments from those related to health care.</p>
4.6	Value is connected to cost	<p>Use this code for comments that reflect beliefs like “You get what you pay for,” that people who are paid better wages will work harder or have more job satisfaction, or any other notion that cost is a predictor of value or quality. Also include comments that reflect the belief that health care providers, corporations, or other parties engage in profiteering.</p> <p>This code is for comments that are not directly related to health and health care; those comments should be coded with the appropriate 1.0 “Role” and/or 3.0 “Decisionmaking” codes. It may be difficult to separate these comments from ones related to health care. Participants who say, for example, “Better hospitals charge more” would be coded here. They are expressing a belief. By contrast, a participant who says, “I wouldn’t go to the less expensive hospital for my care” would be coded at 3.3, “Cost,” because he or she is indicating that cost influences health care choice.</p>
4.7	I am my brother’s keeper	<p>Use this code for comments that reflect the belief that people have a responsibility or moral obligation toward others. This encompasses the belief that one should “do right by others.” For example, a participant might indicate that there’s a moral obligation to help the poor or to support education or to pay taxes or to get vaccinated.</p>

CONTENT DIMENSION 5. Secondary content codes—ALL WILL BE DOUBLE-CODED

Code #	Code name	Code description (inclusion and exclusion)
5.1	Case study	Used to code transcripts with the associated case studies used in the deliberative sessions. <i>This code will be applied administratively when the transcript is uploaded into NVivo. Coders will not use this code.</i>
5.2	Trust	Comments about trusting or not trusting individuals, organizations, governments, institutions, etc. Include comments about individuals/organizations being dishonest, misleading, or having unethical motivations (e.g., “kickbacks” from pharmaceutical companies); also include comments about individuals/organizations actively seeking to protect or inhibit dishonesty. Include comments in which individuals are expressing skepticism about evidence or research findings, guidelines, etc.
5.3	Priorities	Use this code for comments in which participants indicate prioritization or a pecking order that influences their beliefs, reasoning, or conclusions. For example, participants may discuss the “needs of the many” versus the “needs of the few,” or they may indicate situations, circumstances, or philosophies that dictate a priority ordering of choices, limits, or considerations.
5.4	Risk, safety	Comments in which participants explicitly mention risk or safety in connection with health care. Comments may discuss more versus less risk.

PROCESS 6. Secondary process measures—ALL WILL BE DOUBLE-CODED

Code #	Code name	Code description (inclusion and exclusion)
6.1	Reason giving	Use this code to denote chunks of text in which participants are expressing their reasoning for their views.
6.2	Person has changed his/her mind	Comments that provide direct evidence that a person has altered his or her thinking as a result of the conversation or deliberative experience. In coding sections of transcripts where votes are being counted and a participant changes their vote, you may use this code by itself.
6.3	Top-of-mind, preliminary discussion	Used to code to all spontaneous, top-of-mind comments that participants make at the beginning of sessions or case studies, i.e., comments made prior to the facilitator beginning the protocol. <i>This code will be applied administratively when the transcript is uploaded into NVivo. Coders will not use this code.</i>
6.4	Awareness, questioning	Comments indicating that participants were not aware of particular issues, trends, or facts relevant to the discussion. Also code comments indicating that participants question whether “people” are aware or know about these things.

PROCESS 6. Secondary process measures—ALL WILL BE DOUBLE-CODED (continued)

Code #	Code name	Code description (inclusion and exclusion)
6.5	Conversation	Use this code to denote chunks of text in which there is a back-and-forth exchange of short statements focused on the topic at hand, i.e., an interactive conversation among participants in contrast to participants taking turns presenting their points of view.
6.6	Impact of materials on participants	Use this code for comments and conversations that indicate that the prepared materials for the case study have had an impact on participants. For comments expressing awareness issues, use 6.4, “Awareness, questioning” code instead or as well. Use this 6.6 code for other impacts and reactions, when participants indicate a change of mind (also code with 6.2, “Changed mind”), question the validity of the materials, argue with the validity of the underlying research (e.g., researchers studied the wrong people, the studies are old, etc.), engage in wishful thinking (e.g., saying that surely there must be treatments that work), or dismiss evidence-based materials on the basis of personal experience/anecdote (e.g., participants who have had good experience at their local low-volume hospital might dismiss the evidence that high-volume facilities are better overall).
6.7	Impact of experts on participants	Use this code for comments and conversations that indicate that the experts who make presentations or answer questions in a session have had an impact on participants. For comments expressing awareness issues, use the 6.4, “Awareness, questioning” code instead or as well. Use this 6.7 code for other impact and reactions, such as when participants indicate a change of mind (also code with 6.2, “Changed mind”), question the qualifications or expertise of experts, question their motivations (also code for 5.2, “Trust”), etc.
6.8	Wrap-up discussion	Used to code all wrap-up discussion at the end of each session, as applicable (some are short, some extended, and some sessions do not have a wrap-up). <i>This code will be applied administratively when the transcript is uploaded into NVivo. Coders will not use this code.</i>

7. Administrative Codes

Code #	Code name	Code description (inclusion and exclusion)
7.1	Needs Review	Use this only to code sections for which you are uncertain of your coding, so that you can review them later.
7.2	Coder noticed	Information that stands out to the coder but is not captured in other codes (e.g., miscellaneous information).

Appendix F. Telephone Screening Form for Recruiting Participants

In this appendix, we provide the telephone screening form that the recruitment firm in Chicago, IL used as an example of the recruitment process.

AHRQ/AIR Screening Questionnaire for Community Forum Project in Chicago, IL

<input type="checkbox"/> Chicago	<input type="checkbox"/> Wave 1 <input type="checkbox"/> Wave 2 <input type="checkbox"/> Wave 3 <input type="checkbox"/> Wave 4	<input type="checkbox"/> Recruitment ID Number: _____
----------------------------------	--	---

FIRST NAME: _____ LAST NAME: _____ ADDRESS: _____ CITY: _____ STATE: _____ ZIP CODE: _____ COUNTY: _____ PHONE NUMBER(S): DAY: _____ EVENING: _____ E-MAIL 1: _____ E-MAIL 2: _____
--

<i>FOR RECRUITMENT FIRM OFFICE USE:</i>	
ASSIGNED GROUP: _____ SEX: M/F RACE: White/African American/Asian/other _____ EDUCATION: HS grad/GED/technical/associate/1–3 yrs college/4-yr grad/more than 4 yrs CONFIRMATION PHONE CALL? _____ FOLLOWUP PHONE CALL? _____ FOLLOWUP E-MAIL? _____	DATE: _____ TIME: _____ DATE: _____ TIME: _____ DATE: _____ TIME: _____ DATE: _____ TIME: _____ AGE: _____ _____ Hispanic: Yes/No DATE: _____ DATE: _____ DATE: _____

RECRUITMENT INSTRUCTIONS: See Table 1 below with priority population hard quotas that must be met for Chicago. Target numbers are soft quotas.

- Recruit **420** participants who agree to participate in at least two types of discussion groups and are available to participate; **324 must show up and participate in the groups.**
- Recruit participants who have access to the Internet and are familiar with using it.
- Exclusion criteria: Under age 18; currently or formerly employed as a practicing physician, nurse practitioner, registered nurse (RN), licensed practical nurse (LPN), or physician's assistant; recruits who have participated in one study in the past 6 months or two studies in the past year. Do not exclude home health aides.
- Notify AIR staff immediately of any recruitment difficulties, issues with, or questions about the screener.
- Please provide regular recruitment updates (once a day).

Table 1: Priority Population Hard Quotas and Demographic Targets

Total Sample	Chicago 324 Must Show Up	Chicago Recruit 420
Priority Population Hard Quotas (must meet these numbers): Medicare Beneficiary (Aged 65 and over)	33	43
Priority Population Hard Quotas (must meet these numbers): African American Women	61	79
Priority Population Hard Quotas (must meet these numbers): Hispanic	35	45
Target Numbers for Soft Quotas: Female (Sex Demographic)	167	216
Target Numbers for Soft Quotas: Male (Sex Demographic)	157	204
Target Numbers for Soft Quotas: Below 65 Years (Age Demographic)	291	377
Target Numbers for Soft Quotas: 65 Years and Over (Age Demographic)	33	43
Target Numbers for Soft Quotas: Black or African American (Race Demographic)	118	153
Target Numbers for Soft Quotas: White and Other (Race Demographic)	206	267
Target Numbers for Soft Quotas: Hispanic or Latino (Hispanic or Latino Demographic)	84	109
Target Numbers for Soft Quotas: Non-Hispanic (Hispanic or Latino Demographic)	240	311
Target Numbers for Soft Quotas: Urban (Urban Demographic)	324	420
Target Numbers for Soft Quotas: High School Grad or Less* (Education Demographic)	137	176

**AIR recognizes that [recruitment firm] recommends a lower Target for High School Grad or Less. The above numbers are reflective of Census numbers (42 percent). AIR requests that [recruitment firm] strive to recruit 20 percent if 42 percent is not attainable.*

Table 2: Five Recruitment Groups

Details about the Five Recruitment Groups	Number of Participants and Groups	General Population Show	General Population Recruit
Brief Citizens' Deliberation	Number of participants per group	12	16
Brief Citizens' Deliberation: 2-hour in-person session	Number of groups	6	6
Brief Citizens' Deliberation: \$100 incentive	Brief Citizens' Deliberation total sample	72	96
Online Deliberative Polling[®]	Number of participants per group	NA	NA
Online Deliberative Polling[®]: (Online sessions) four 1.25-hour sessions to occur over 4 weeks	Number of groups	24	24
Online Deliberative Polling[®]: \$125 incentive	Online Deliberative Polling [®] total sample	72	96
Citizens' Panel	Number of participants per group	24	30
Citizens' Panel: One Group: 20-hour session over 3 days (8 hours Day 1, 8 hours Day 2, 4 hours Day 3)	Number of groups	1	1
Citizens' Panel: \$325 incentive	Citizens' Panel total sample	24	30
Community Deliberation	Number of participants per group	12	16
Community Deliberation: 2.5-hour in-person session/2.5-hour in-person session the following week	Number of groups	6	6
Community Deliberation: \$160 incentive	Community Deliberation total sample	72	96
Deliberative Groups Subtotal	Number of groups	37	37
	Deliberative Groups total	240	318
Reading Materials Only Group \$75 incentive	Reading Materials Only Group total sample	84	102
	Recruitment groups total sample	324	420

IF RESPONDENT IS NOT AVAILABLE, EITHER:

- Determine a good time to call back, **OR**
- Leave message on answering machine

VOICEMAIL SCRIPT: Hello, my name is [*RECRUITER'S FIRST AND LAST NAME*]. I'm calling on behalf of the American Institutes for Research (AIR), a nonprofit research organization. We are calling to invite you to take part in community discussion groups about health care quality. Please call [*INSERT PHONE NUMBER*] at your convenience.

INTRODUCTION WHEN PERSON ANSWERS THE PHONE

Hello, may I please speak with [*FIRST AND LAST NAME*]? My name is [*RECRUITER'S FIRST AND LAST NAME*], and I'm calling on behalf of the American Institutes for Research (AIR), a nonprofit research organization.

IF RESPONDENT IS AVAILABLE, INTRODUCE SELF AND REASON FOR CALLING.

I'm calling today to tell you about an exciting opportunity to be involved in discussion groups taking place in your community about health care quality. Leading health care experts and Government decisionmakers want to hear from members of the community like you. These groups are a chance for you to share your opinions.

Four different types of discussion groups will be taking place in your community at [location address]. If you are interested in voicing your opinion about important health care issues, and you meet study requirements, we will invite you to participate in one of these groups this summer or fall. You would be paid for taking part.

May I ask you a few questions? *IF YES, CONTINUE TO Q1. IF NO, THANK & END (END SCRIPT B).*

IF NEEDED, EXPLAIN FURTHER: Because we want to include people who are a mix of different backgrounds and experiences, I need to ask you a few questions to see if you qualify.

ELIGIBILITY CRITERIA QUESTIONS

[ENTER RESPONSES INTO DATABASE. **ALL QUESTIONS MUST HAVE A RESPONSE ENTERED** EVEN IF THE QUESTIONS ARE SKIPPED OR THE ANSWER IS NO.]

Q1. Part of this project involves taking surveys online. Would you be able to take them online?

☐ YES

☐ NO, *END SCRIPT A*

Q2. How comfortable do you feel completing a survey online?

☐ Not very comfortable, *END SCRIPT A*

☐ Somewhat comfortable

☐ Very comfortable

Q3. Are you comfortable reading and having conversations in English?

☐ YES

☐ NO, *THANK & END (END SCRIPT A)*

Q4. How old were you on your last birthday? (*NONE UNDER 18; RECRUIT MIX*)

IF UNDER 18, THANK & END (*END SCRIPT A*)

Q5. Are you or have you ever been employed as a practicing physician, nurse practitioner, registered nurse (RN), licensed practical nurse (LPN), or physician's assistant?

☐ YES, *END SCRIPT A*

☐ NO

Q6. Have you had any other job in the health care field?

☐ NO

☐ YES, (WRITE IN OCCUPATION) _____

Q7. What is your gender? (*RECRUIT MIX; SEE TABLE 1*)

☐ MALE

☐ FEMALE

- Q8. What is the highest grade or level of school you have completed?
(*RECRUIT MIX: 42 percent High School Grad or Less*)
- ☐ Less than high school graduate
 - ☐ High school graduate
 - ☐ GED
 - ☐ Technical or vocational school or certificate program
 - ☐ Associate's degree (2-year college graduate)
 - ☐ 1 or more years of college but didn't graduate
 - ☐ 4-year college graduate
 - ☐ More than 4-year college graduate
- Q9. Are you of Hispanic, Latino, or Spanish background?
- ☐ YES
 - ☐ NO
- Q10. How would you describe your race? (SEE TABLE 1, PRIORITY POPULATION QUOTAS)
CHECK ALL THAT APPLY
- ☐ American Indian or Alaska Native
 - ☐ Asian or Pacific Islander
 - ☐ Black or African American
 - ☐ White
 - ☐ Another race (write in) _____
- Q11. Do you speak a language other than English at home?
- ☐ YES
 - ☐ NO, *SKIP TO Q13*
- Q12. What is this language? (RECRUIT 45 BILINGUAL HISPANICS)
- ☐ SPANISH
 - ☐ ANOTHER LANGUAGE
- Q13. Finally, if you agree to take part in the discussion groups, we will need to provide your e-mail address and phone number to AIR so that they can contact you to provide the details about taking part in the groups. Will that be okay with you?
- ☐ YES
 - ☐ NO, *END SCRIPT A*
 - ☐ MAYBE, BUT I WOULD LIKE MORE INFORMATION BEFORE DECIDING
→(*Provide name and phone number where participants can get more information. Make arrangements to call back.*)

For more information about this project, contact [contact information].

IF INDIVIDUAL MEETS ELIBILITY CRITERIA:

Thank you for answering all of my questions. It looks like you're eligible to participate in the community discussion! Four different types of discussion groups will be held in your community this summer and fall. Some groups are longer, some are shorter, and some happen online, although most of them are in person. All the groups will give participants the opportunity to discuss important health care issues with people from your community.

I'm going to read you a description of each type of group and ask you to tell me which ones you'd be interested in. Then we'll go through the dates and times of the groups. Once we see your availability, AIR will assign you either to one of the discussion groups or to a materials-only group. The assignment will be made "at random," which means by chance, like flipping a coin. If you are assigned to the reading materials group, we will ask you to read educational materials at home but not take part in the discussion groups. Everyone assigned to a group will be asked to complete surveys before and after the groups take place. If your preferred groups are full, you might not be selected to take part at this time.

You may want to get your calendar to check your availability.

Table 3: Discussion Group Interest and Availability

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none">• Meet once, in person, for 2 hours to discuss important issues related to the quality of health care• Receive \$100 for participating and completing three surveys	101	Tuesday, Aug. 14, from 12:30 to 3:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none">• Meet once, in person, for 2 hours to discuss important issues related to the quality of health care• Receive \$100 for participating and completing three surveys	102	Tuesday, Aug. 14, from 12:30 to 3:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none">• Meet once, in person, for 2 hours to discuss important issues related to the quality of health care• Receive \$100 for participating and completing three surveys	103	Wednesday, Aug. 15, from 12:30 to 3:00 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none"> • Meet once, in person, for 2 hours to discuss important issues related to the quality of health care • Receive \$100 for participating and completing three surveys 	104	Wednesday, Aug. 15, from 5:30 to 8:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none"> • Meet once, in person, for 2 hours to discuss important issues related to the quality of health care • Receive \$100 for participating and completing three surveys 	105	Thursday, Aug. 16, from 12:30 to 3:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Brief Citizens' Deliberation (BCD), you will: <ul style="list-style-type: none"> • Meet once, in person, for 2 hours to discuss important issues related to the quality of health care • Receive \$100 for participating and completing three surveys 	106	Thursday, Aug. 16, from 5:30 to 8:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Citizens' Panel (CP), you will: <ul style="list-style-type: none"> • Listen to presentations from leading experts in the field about important issues related to the quality of health care • Question the experts about health care • Meet for 2½ days, from Friday to Sunday afternoon • Receive \$325 for participating and completing three surveys 	201	Friday, Aug. 17, from 9:00 a.m. to 5:00 p.m. CT, & Saturday, Aug. 18, from 9:00 a.m. to 5:00 p.m. CT & Sunday, Aug. 19, from 9:00 a.m. to 1:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	301	Tuesday, Aug. 14, from 12:00 to 3:00 p.m. CT & Tuesday, Aug. 21, from 12:00 to 3:00 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	302	Tuesday, Aug. 14, from 6:00 to 9:00 p.m. CT & Tuesday, Aug. 21, from 6:00 to 9:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	303	Wednesday, Aug. 15, from 12:00 to 3:00 p.m. CT & Wednesday, Aug. 22, from 12:00 to 3:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	304	Wednesday, Aug. 15, from 6:00 to 9:00 p.m. CT & Wednesday, Aug. 22, from 6:00 to 9:00 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	305	Thursday, Aug. 16, from 12:00 to 3:00 p.m. CT & Thursday, Aug. 23, from 12:00 to 3:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Community Deliberation (CD), you will: <ul style="list-style-type: none"> • Meet in person for 2.5 hours to discuss important issues related to the quality of health care and then again in person for 2.5 hours a week later • Interact with other group members and with health care experts online for the week between the two in-person meetings • Receive \$160 for participating and completing three surveys 	306	Thursday, Aug. 16, from 6:00 to 9:00 p.m. CT & Thursday, Aug. 23, from 6:00 to 9:00 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®] (ODP[®]), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	401	Tuesday, Oct. 2, from 6:00 to 7:15 p.m. CT & Tuesday, Oct. 9, from 6:00 to 7:15 p.m. CT & Tuesday, Oct. 16, from 6:00 to 7:15 p.m. CT & Tuesday, Oct. 23, from 6:00 to 7:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®] (ODP[®]), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	402	Tuesday, Oct. 2, from 7:00 to 8:15 p.m. CT & Tuesday, Oct. 9, from 7:00 to 8:15 p.m. CT & Tuesday, Oct. 16, from 7:00 to 8:15 p.m. CT & Tuesday, Oct. 23, from 7:00 to 8:15 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Online Deliberative Polling[®], you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	403	Tuesday, Oct. 2, from 8:00 to 9:15 p.m. CT & Tuesday, Oct. 9, from 8:00 to 9:15 p.m. CT & Tuesday, Oct. 16, from 8:00 to 9:15 p.m. CT & Tuesday, Oct. 23, from 8:00 to 9:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®], you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	404	Tuesday, Oct. 2, from 9:00 to 10:15 p.m. CT & Tuesday, Oct. 9, from 9:00 to 10:15 p.m. CT & Tuesday, Oct. 16, from 9:00 to 10:15 p.m. CT & Tuesday, Oct. 23, CT from 9:00 to 10:15 p.m. ET	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®], you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	405	Wednesday, Oct. 3, from 6:00 to 7:15 p.m. CT & Wednesday, Oct. 10, from 6:00 to 7:15 p.m. CT & Wednesday, Oct. 17, from 6:00 to 7:15 p.m. CT & Wednesday, Oct. 24, from 6:00 to 7:15 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Online Deliberative Polling[®] (ODP[®]), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	406	Wednesday, Oct. 3, from 7:00 to 8:15 p.m. CT & Wednesday, Oct. 10, from 7:00 to 8:15 p.m. CT & Wednesday, Oct. 17, from 7:00 to 8:15 p.m. CT & Wednesday, Oct. 24, from 7:00 to 8:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®] (ODP[®]), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	407	Wednesday, Oct. 3, from 8:00 to 9:15 p.m. CT & Wednesday, Oct. 10, from 8:00 to 9:15 p.m. CT & Wednesday, Oct. 17, from 8:00 to 9:15 p.m. CT & Wednesday, Oct. 24, from 8:00 to 9:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling[®] (ODP[®]), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	408	Wednesday, Oct. 3, from 9:00 to 10:15 p.m. CT & Wednesday, Oct. 10, from 9:00 to 10:15 p.m. CT & Wednesday, Oct. 17, from 9:00 to 10:15 p.m. CT & Wednesday, Oct. 24, from 9:00 to 10:15 p.m. CT	<input type="checkbox"/>

CHECK IF INTERESTED METHOD	DISCUSSION GROUP DESCRIPTION	CODE	DATES AND TIMES	CHECK IF AVAILABLE
<input type="checkbox"/>	For Online Deliberative Polling® (ODP®), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	409	Thursday, Oct. 4, from 6:00 to 7:15 p.m. CT & Thursday, Oct. 11, from 6:00 to 7:15 p.m. CT & Thursday, Oct. 18, from 6:00 to 7:15 p.m. CT & Thursday, Oct. 25, from 6:00 to 7:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling® (ODP®), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	410	Thursday, Oct. 4, from 7:00 to 8:15 p.m. CT & Thursday, Oct. 11, from 7:00 to 8:15 p.m. CT & Thursday, Oct. 18, from 7:00 to 8:15 p.m. CT & Thursday, Oct. 25, from 7:00 to 8:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling® (ODP®), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	411	Thursday, Oct. 4, from 8:00 to 9:15 p.m. CT & Thursday, Oct. 11, from 8:00 to 9:15 p.m. CT & Thursday, Oct. 18, from 8:00 to 9:15 p.m. CT & Thursday, Oct. 25, from 8:00 to 9:15 p.m. CT	<input type="checkbox"/>
<input type="checkbox"/>	For Online Deliberative Polling® (ODP®), you will: <ul style="list-style-type: none"> • Discuss important health care issues with participants from around the United States • Question leading health care experts online • Meet online for 1 hour and 15 minutes once a week for 4 weeks • Receive \$125 for participating and completing three surveys 	412	Thursday, Oct. 4, from 9:00 to 10:15 p.m. CT & Thursday, Oct. 11, from 9:00 to 10:15 p.m. CT & Thursday, Oct. 18, from 9:00 to 10:15 p.m. CT & Thursday, Oct. 25, from 9:00 to 10:15 p.m. CT	<input type="checkbox"/>

Reading Materials Group

If you are assigned to the group to receive materials to read at home about important health care issues, you will be paid \$75 for reading the materials accessible online and taking two surveys.

NOTE: RECRUITED PARTICPANTS *MUST AGREE TO PARTICIPATE IN AT LEAST TWO GROUPS*. IF THE PERSON SELECTS ONLY ONE GROUP, SAY:

AIR is interested in learning about the types of discussion groups that work better and are preferred by more people. Because they'll be comparing the groups to each other, in order to be included, you'll need to select two or more types of discussion groups.

AFTER TWO GROUPS ARE SELECTED:

Thank you for indicating your availability. We will call you to let you know if you are selected for a group, as well as the dates, times, locations, and other details.

You will receive an e-mail with a link to complete an online survey. Please complete your survey as soon as you get it. This e-mail will come from [e-mail address]. Please add this e-mail address as well to your e-mail contact list to prevent this e-mail from being blocked by your e-mail account's SPAM blocker.

Here are a few more details:

Table 4: Discussion Group Details

Additional Details About the Discussion Groups	INITIALS OF RECRUITER
The project is funded by the Agency for Healthcare Research and Quality (AHRQ), which is part of the U.S. department of Health and Human Services.	_____
The discussion groups are part of a research project to get input from people about improving health care quality and to learn more about the best ways to get public input on important health care issues. There are no known risks to taking part in this type of study. Although there are no direct benefits to participants, what we learn will help policymakers make more informed decisions on health care policy issues, which could improve the quality of health care.	_____
The information you shared with me today, as well as any information you provide on the surveys will not be shared with anyone outside the AIR project team in a way that could identify you.	_____
The discussion groups will be video- and audiotaped to help the researchers understand what was said in the group discussions. The recordings will be used for research and staff training and may be used in presentations of the project's findings.	_____
Your name will not be used in any report or publication, although the video recordings of the in-person discussion groups may be viewed by research and training staff at AIR as well as by external audiences.	_____

If you agree to take part, I need to verify the spelling of your name, mailing address, and e-mail address. We will not use this information for any purpose outside of this study. **CONFIRM RESPONDENT'S INFORMATION.**

Name: _____ Telephone: _____

Address: _____

City, State: _____ ZIP: _____

E-MAIL 1: _____

E-MAIL 2: _____

IF NO E-MAIL ADDRESS, ASK: Would you be willing to create a free e-mail account through Gmail or Yahoo for this project?

Also, please let me confirm that I have all your correct phone numbers so that we may call and confirm with you. **CONFIRM RESPONDENT'S INFORMATION.**

Home phone: _____ Other phone: _____

If you have any questions or find that you can't attend, please call us right away at [insert phone number]. Thank you for your time and for agreeing to help.

If you have questions or would like more information about this project, contact [contact information]

If you have concerns or questions about your rights as a participant, contact [contact information].

END SCRIPT A (FOR PEOPLE WHO DO NOT MEET SELECTION CRITERIA):

[Recruitment firms to use standard script for noneligible participants.]

END SCRIPT B (GENERIC THANK & END):

I appreciate your taking the time to speak with me, and I hope you have a good day.

Appendix G. Informed Consent Form for In-Person Discussion Groups

What is this project about and what will you ask me to do?

The Community Forum project will hold several discussion groups around the U.S. to hear from community members about what they think and feel about using research to improve health care. The purpose of the discussion group that you are taking part in is to help us improve our methods for carrying out the discussion groups and our procedures for collecting the data that will be used to see how well the discussion groups work.

After taking part in the discussion group, you will be asked to complete a short survey about the topics discussed and your experience in the discussion groups.

There are no known risks to taking part in this type of study. Although there are no direct benefits to participants, what we learn will help policymakers make more informed decisions on health care policy issues, which could improve the quality of health care.

Who is doing this project?

This research project is being done by the American Institutes for Research (AIR), a non-profit research organization. The project is paid for by the Agency for Healthcare Research and Quality, a government agency.

Do I have to participate in this project?

No. It is your choice to participate or not. Also, you can stop participating at any time, and you do not have to answer any questions that you do not want to. If you choose not to participate or stop participating, there are no penalties. All information you provide today is confidential and will not be linked to you as an individual.

Will you be recording the discussions?

Yes. The discussion groups will be video and audio-taped to help the researchers understand what was said in the group discussions. The recordings will be used for research, training staff, and may be used in presentations of the project's findings.

How will you protect my privacy?

We will not use your name in any report or publication, although the video recordings of the in-person discussion groups may be viewed by research and training staff at AIR as well as external audiences.

What if I want more information?

If you want more information about this project, please contact [contact information].

If you have concerns or questions about your rights as a participant, you can contact the Chair of AIR's Institutional Review Board at [email address], or toll free at [phone number] or the Project Director [contact information].

Please sign below if you agree to participate.

Signing your name below means that you are giving your “informed consent” to participate today. This means that you have read and understood the information on this form, you have had a chance to ask questions, and you are willing to participate under the conditions we have described. Your signature does NOT allow AIR or AHRQ to identify you by name when your comments are used.

Your signature: _____ Today's date: _____

Please print your name: _____

Appendix H. Technical Description of Randomization and Weighting Procedures

Below, we further elaborate the original sample allocation plan, the actual sample allocation achieved in the full study, the randomization scheme, and how the randomization scheme was incorporated into the quantitative analyses to answer the research questions.

Sample Allocation

In this study, there were five experimental conditions, including four deliberative methods and a control condition. Control group participants were asked to read educational materials on the deliberative issues, but were not convened in groups to deliberate.

Exhibit H1 shows the planned sample allocation across the four geographic locations and within each of the five experimental conditions. We planned to recruit an equal number of participants across the four locations (e.g., 420 recruits with 324 to show within each location). It is important to note that for the Online Deliberative Polling[®] (ODP[®]) method, we planned to construct mixed-location groups. That is, each online group would include participants coming from all four geographic locations, and the location composition (i.e., the proportion of participants coming from each location within a group) could vary slightly from group to group.

Exhibit H1. Planned sample allocation across locations and experimental conditions

Deliberative Method	Number of Participants and Groups	Chicago, IL: show	Chicago, IL: recruit	Durham, NC: show	Durham, NC: recruit	Silver Spring, MD: show	Silver Spring, MD: recruit	Sacramento, CA: show	Sacramento, CA: recruit	Overall: show	Overall: recruit
Brief Citizens' Deliberation (Method 1)	# participants per group	12	16	12	16	12	16	12	16	12	16
	# groups	6	6	6	6	6	6	6	6	24	24
	Total sample	72	96	72	96	72	96	72	96	288	384
Citizens' Panel (Method 2)	# participants per group	24	30	24	30	24	30	24	30	24	30
	# groups	1	1	1	1	1	1	1	1	4	4
	Total sample	24	30	24	30	24	30	24	30	96	120
Community Deliberation (Method 3)	# participants per group	12	16	12	16	12	16	12	16	12	16
	# groups	6	6	6	6	6	6	6	6	24	24
	Total sample	72	96	72	96	72	96	72	96	288	384
Online Deliberative Polling® (Method 4)	# participants per group	12	16	12	16	12	16	12	16	12	16
	# groups	NA	NA	NA	NA	NA	NA	NA	NA	24	24
	Total sample	72	96	72	96	72	96	72	96	288	384
Deliberative Groups Subtotal	# groups	NA	NA	NA	NA	NA	NA	NA	NA	76	76
	Total	240	318	240	318	240	318	240	318	960	1272
Control (Method 0)	Total sample	84	102	84	102	84	102	84	102	336	408
Total Deliberation plus Control	# groups	NA	NA	NA	NA	NA	NA	NA	NA	76	76
	Total sample	324	420	324	420	324	420	324	420	1296	1680

Our original recruitment plan took into account a 30 percent no-show rate. To be more specific, for the methods of Brief Citizens' Deliberation (BCD), Community Deliberation (CD), and ODP, we planned to recruit 16 people for each group and expected 12 people to attend the deliberative sessions. For the Citizens' Panel (CP) method, we planned to recruit 30 people and expected 24 to attend.

During the recruitment process, however, we found that the show rates were lower than our expectation for the CP, CD, and ODP methods, and some groups did not have sufficient numbers of Hispanic participants to form a heterogeneous group of participants from diversified backgrounds. Therefore, we increased the target group size for the CP and CD methods to 32 and 18, respectively, in Silver Spring, MD, and Sacramento, CA. In addition, since the recruitment response rate was lower in Sacramento than in the other three locations, we recruited 39 additional people in Sacramento and randomized them across different experimental conditions to ensure an adequate number of participants showed. We also directly recruited 27 participants in Silver Spring and Sacramento to ensure an adequate number of participants showed for a particular method (e.g., BCD, CP, or CD methods) and directly recruited Hispanic participants to increase group diversity in the CP groups.

Exhibits H2 and H3 display (1) the number of participants expected to show, (2) the number of people who actually showed, (3) the planned number of recruits, and (4) the actual number of recruits, within each of the five experimental conditions and within each of four geographic locations, respectively. A person is considered a valid participant (i.e., an actual show) for the quantitative analyses if he/she attended the deliberation sessions that he/she was assigned to and completed both the pre-survey and post-survey.

Exhibit H2. Planned and actual sample size for recruitment and attendance across experimental conditions

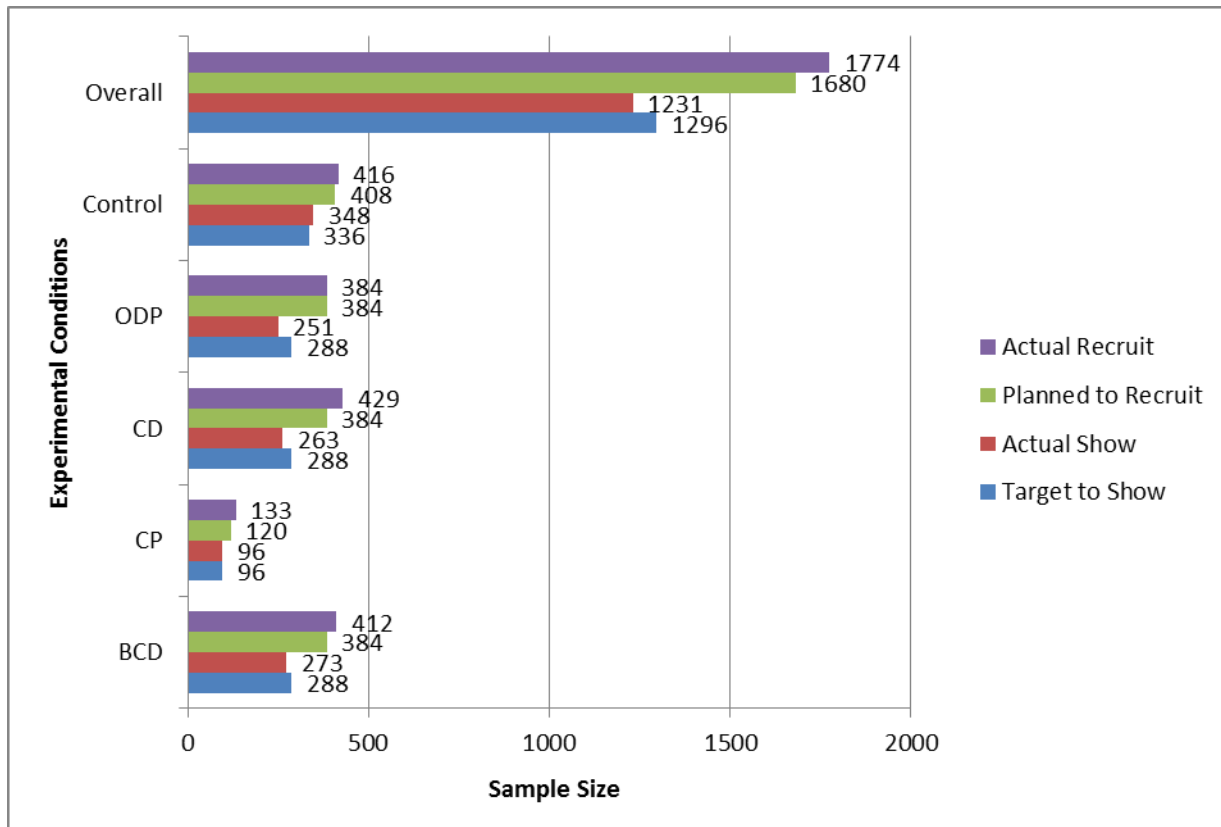
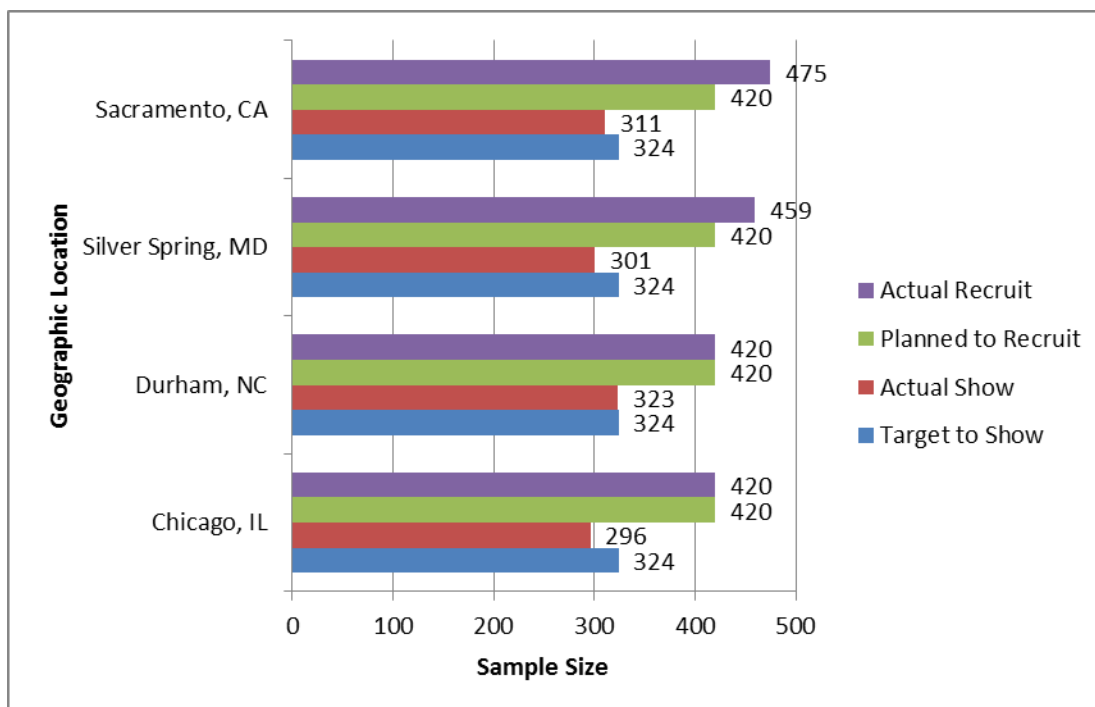


Exhibit H3. Planned and actual sample size for recruitment and attendance across geographic locations



Randomization Scheme

Within each of the four geographic locations, selected participants were randomly assigned to the five experimental conditions (four deliberative methods or control). In a multi-treatment randomized study, it is typical to recruit a sample of participants who are willing and able to participate in any of the treatment options and then randomize the participants across those options and the control condition. This approach is consistent with maximizing internal validity, a priority for this study.

But, for this study, a major limitation of the typical randomization approach is the disparate burden placed on participants by the deliberative methods, since the demands of the methods vary considerably. This difference introduces the potential for selection bias—participants who were able and willing to participate in lower-burden or online methods might differ significantly in their personal characteristics (e.g., age) or in some unobserved ways from those who were able and willing to participate in higher-burden or in-person methods. Because of the variation in burden and deliberation mode, the pool of individuals who would find all four methods acceptable (and thus be available for randomization using the typical approach) would most likely be small and not representative of the populations we hoped to represent (see Chapter 3). Thus, the typical randomization approach would likely be extremely expensive and/or have poor external validity.

One alternative was to draw a separate random sample for each method, each with its own control group. However, because of the variation in demands of the methods, the resulting samples would likely differ substantially from each other, threatening internal validity. Because of the inherent selection bias, this approach would limit our ability to compare specific deliberative methods to others.

To address this dilemma, we used a two-step randomization procedure designed to maximize the pool of potential willing participants, minimize the cost of recruitment, and yield internally and externally valid comparisons among specific deliberative methods. We made these relevant comparisons among the methods using only the subset of sample members who were willing to accept assignment to the methods being compared to each other. This approach minimizes selection bias in each comparison. After the randomization process was completed, an assignment dataset and a randomization dataset were created. The former includes the method/group assignment after randomization, and the latter includes all the individual-level information and results we obtain from the randomization, such as each participant's preferences and availability across the four deliberative methods, and the probability of being assigned to each experimental condition. The details of all the variables in these datasets will be discussed in the "Assignment Dataset" and "Randomization Dataset" sections.

Screening Data

Potential participants from the recruitment firms were provided descriptions of the five experimental conditions and asked to indicate, on the electronic form or on the phone, which of the four deliberative methods they were willing to participate in. They were asked to select as many of the methods as they were interested in and would take part in. The instructions further explained that they would be randomly assigned to one of their selected methods or a control condition, where they would be asked to read materials but not convene or take part in discussions. Subsequently, for each of the methods in which they showed interest, a set of time slots were given, and they were asked to select all the time slots for which they were available.

They were required to select (i.e., be both willing and available to participate in) at least two of the four methods to be eligible for the study. It was necessary to exclude subjects who only selected one method because we could not use their outcomes to compare methods.

Exhibit H4 presents an example of the method preference information we collected from an eligible participant. Four binary indicators were used to identify each person's willingness to participate in the four deliberative methods (MTH1_BCD, MTH2_CP, MTH3_CD, MTH4_ODP) with

1=interested and 0=not interested. In addition, a set of binary indicators '101' to '412' were used to identify a person's availability for each time slot provided, with 1=available and 0=not available:

- 101–106 are indicators for the six time slots of Method 1 (BCD);
- 201 is an indicator for the one single time slot of Method 2 (CP);
- 301–306 are indicators for the six time slots of Method 3 (CD);
- 401–412 are indicators for the twelve time slots of Method 4 (ODP).

Note that although there were a total of 24 ODP groups, we held two groups concurrently within each time slot; thus, each participant could choose from 12 time slots.

Exhibit H4. Example participant interest and availability

MTH1_BCD	MTH2_CP	MTH3_CD	MTH4_ODP	101	102	103	104	105	106	201	301	...	401	...
1	1	1	0	0	0	1	1	1	1	0	1	...	0	...

In the above example, the person was willing to participate in Methods 1, 2, and 3 (i.e., BCD, CP, and CD), but only available to participate in Methods 1 and 3. Therefore, in the randomization dataset we created a set of indicators to indicate participants' willingness *and* availability to join each method (i.e., variables 'PREF0_1' to 'PREF0_4'). If a person indicated his/her interest in participating in a method *M* and was available for *at least one* of the provided time slots for the method, then PREF0_M=1, otherwise PREF0_M=0.

Randomization Procedure

To inform participants about their method and group assignment as soon as possible, the randomization was performed on a weekly basis. After receiving the screening data of recruits from the previous week, we randomly assigned recruits to one of the five experimental conditions. For those assigned to a deliberative method, we then non-randomly assigned them to the deliberative group with the fewest participants at that moment, within that method. The assignment procedure continued until the target sample sizes were reached for all the deliberative groups and all the experimental conditions. The target sample sizes across conditions and locations are listed in Exhibit H5. Note that the target group size for the CP and CD methods was increased to 32 per group and 18 per group in Silver Spring and Sacramento to ensure an adequate number of participants who would show up. Also, an additional wave of recruitment and randomization was conducted in Sacramento (n=39) due to the high cancellation rate after the initial recruitment. These details will be discussed in the "Random Assignment for the Last Wave of Sacramento Data" section. Below, we will introduce the assignment process in detail,

including the two-step randomization process to deliberative methods, as well as the non-random assignment to deliberative groups within a method.

Exhibit H5. Target sample sizes for assignment across experimental conditions and geographic locations

Location	Method 1 (BCD)	Method 2 (CP)	Method 3 (CD)	Method 4 (ODP)	Method 0 (Control)
Chicago, IL	96 total 16 per group	30 total 30 per group	96 total 16 per group	96 total 16 per group	102 total
Durham, NC	96 total 16 per group	30 total 30 per group	96 total 16 per group	96 total 16 per group	102 total
Silver Spring, MD	96 total 16 per group	32 total 32 per group	108 total 18 per group	96 total 16 per group	102 total
Sacramento, CA	113 total 16 per group	32 total 32 per group	122 total 18 per group	96 total 16 per group	110 total

Step 1 – Randomization to Control vs. Deliberation

It was assumed that all the participants were available to take part in the control group, regardless of how many or which deliberative methods they selected. Therefore, in the randomization process, we first randomly assigned each participant to the control group with a fixed probability. In Chicago, which is the first geographic location where we started recruitment, we decided to set a conservative criterion—each participant had a 10 percent probability to be randomly assigned to the control group. For the other three locations, a more lenient criterion was used and each participant had a 15 percent probability to be assigned to the control group. Of note, if all the deliberative methods a participant selected reached the target sample size, he/she was assigned to the control group with 100 percent probability.

Step 2 – Randomization to a Deliberative Method

If a participant was not assigned to the control group in Step 1, then we examined both the preference information from the participant and the current assignment information for that location to find a set of “open” methods that:

1. the participant was willing and available to participate in (i.e., PREF0=1) and
2. had open slots that the participant is available for.

In the randomization dataset, we created a set of indicators, PREF1_1 to PREF1_4, to indicate the “open” methods that satisfy both of the above criteria. Among all the “open” methods (i.e., PREF1=1), we conducted the random assignment with equal probability of assignment to each method. As illustrated in Exhibit H6, this procedure established up to 16 pools of participants with different probabilities of assignment to each experimental condition. In the first column of Exhibit H6, we identified all possible combinations of the “open” methods where A through D represents the four methods and X represents the control group. Here we take as an example a 15 percent probability of assignment to the control group (i.e., for Durham, Silver Spring, and Sacramento). The assignment probabilities assuming a 10 percent probability of assignment to the control group are omitted in this document, but can be calculated in a similar manner. The assignment probabilities for each individual participant in our sample are recorded in the randomization dataset with variable names ‘PROB_ASSIGN_0’ to ‘PROB_ASSIGN_4’.

Step 3 – Non-Random Assignment to a Deliberative Group

After a participant was assigned to a particular method, we looked for all the deliberative groups, within that method, that he/she was available to participate in and that were not full. We then checked the number of recruits already assigned to each available group and assigned the participant to the group with the fewest participants at that moment. This strategy allowed us to make group assignments efficiently within each method. Since our research objective was to compare the effectiveness of deliberation across different methods but not at the deliberative group level, the non-random assignment to group within each method did not threaten the internal validity.

Exhibit H6. Potential combinations of open methods and the corresponding assignment probabilities

Combination of “Open” Methods	Probability of Assigning to Each Experimental Condition: Method 1 (BCD)	Probability of Assigning to Each Experimental Condition: Method 2 (CP)	Probability of Assigning to Each Experimental Condition: Method 3 (CD)	Probability of Assigning to Each Experimental Condition: Method 4 (ODP)	Probability of Assigning to Each Experimental Condition: Method 0 (Control)
ABCDX	21.25%	21.25%	21.25%	21.25%	15.00%
ABCX	28.33%	28.33%	28.33%		15.00%
ABDX	28.33%	28.33%		28.33%	15.00%
ACDX	28.33%		28.33%	28.33%	15.00%
BCDX		28.33%	28.33%	28.33%	15.00%
ABX	42.50%	42.50%			15.00%
ACX	42.50%		42.50%		15.00%
ADX	42.50%			42.50%	15.00%
BCX		42.50%	42.50%		15.00%
BDX		42.50%		42.50%	15.00%
CDX			42.50%	42.50%	15.00%
AX	85.00%				15.00%
BX		85.00%			15.00%
CX			85.00%		15.00%
DX				85.00%	15.00%
X					100.00%

Random Assignment for the Last Wave of Sacramento, CA, Data

After the initial four waves of recruitment in Sacramento, we found that quite a few recruits cancelled their participation before the deliberation started for the BCD and CD methods. To ensure we had adequate number of participants in each group, we decided to recruit another 40 people to fill in the slots due to cancellations. The same randomization process was conducted for these 40 people as for the other recruits, except that the group sizes were specified based on the number of cancellations in each group:

The target group sizes for the CP and ODP methods were set as 0 because no additional recruits were needed for those methods at that time. The probability of being assigned to the control group was set as 15 percent, but after all the slots were filled for BCD and CD, the rest of the recruits were assigned to the control group with 100 percent probability.

Assignment Dataset

After the group assignment was complete for a wave of recruitment, we created an assignment dataset including information of each recruit's method and group assignment. The coding for the methods and groups is listed in Exhibit H7.

Exhibit H7. Coding of experimental conditions and deliberative groups across all locations

Method Name	Location	Value Coding for Variables in the Assignment Dataset: Method	Value Coding for Variables in the Assignment Dataset: Group76
BCD	Chicago	1	1 to 6
BCD	Durham	1	7 to 12
BCD	Silver Spring	1	13 to 18
BCD	Sacramento	1	19 to 24
CP	Chicago	2	25
CP	Durham	2	26
CP	Silver Spring	2	27
CP	Sacramento	2	28
CD	Chicago	3	29 to 34
CD	Durham	3	35 to 40
CD	Silver Spring	3	41 to 46
CD	Sacramento	3	47 to 52
ODP	Mixed Location	4	53 to 76
Control	Chicago	0	77
Control	Durham	0	78
Control	Silver Spring	0	79
Control	Sacramento	0	80

The assignment data was then incorporated into our master database and the information was sent back to recruitment firms for follow-up and participation confirmation.

Randomization Dataset

At the completion of randomization process, we created a randomization dataset with information about all of the recruits, except for those directly recruited. A summary of all the variables included in the randomization dataset is listed in Exhibit H8. As elaborated in the later section, this dataset was used to incorporate the randomization scheme into the quantitative analyses when making comparisons across deliberative methods.

Exhibit H8. Variable names, labels, and possible values in the randomization dataset

Variable Names	Variable Labels	Possible Values
PREF0_1	Willing and available to participate in at least one time slot of Method 1 (BCD)	1=yes, 0=no
PREF0_2	Willing and available to participate in at least one time slot of Method 2 (CP)	1=yes, 0=no
PREF0_3	Willing and available to participate in at least one time slot of Method 3 (CD)	1=yes, 0=no
PREF0_4	Willing and available to participate in at least one time slot of Method 4 (ODP)	1=yes, 0=no
PREF1_1	Willing and available to participate, slots open for Method 1 (BCD)	1=yes, 0=no
PREF1_2	Willing and available to participate, slots open for Method 2 (CP)	1=yes, 0=no
PREF1_3	Willing and available to participate, slots open for Method 3 (CD)	1=yes, 0=no
PREF1_4	Willing and available to participate, slots open for Method 4 (ODP)	1=yes, 0=no
CONTRAST0_1	Willing and available to participate in both Method 1 (BCD) and Method 2 (CP)	1=yes, 0=no
CONTRAST0_2	Willing and available to participate in both Method 1 (BCD) and Method 4(ODP)	1=yes, 0=no
CONTRAST0_3	Willing and available to participate in both Method 2 (CP) and Method 3 (CD)	1=yes, 0=no
CONTRAST0_4	Willing and available to participate in both Method 1 (BCD) and Method 3 (CD)	1=yes, 0=no
CONTRAST0_5	Willing and available to participate in both Method 2 (CP) and Method 4 (ODP)	1=yes, 0=no
CONTRAST0_6	Willing and available to participate in both Method 3 (CD) and Method 4 (ODP)	1=yes, 0=no
CONTRAST1_1	Willing and available to participate, slots open for both Method 1 (BCD) and Method 2 (CP)	1=yes, 0=no
CONTRAST1_2	Willing and available to participate, slots open for both Method 1 (BCD) and Method 4(ODP)	1=yes, 0=no
CONTRAST1_3	Willing and available to participate, slots open for both Method 2 (CP) and Method 3 (CD)	1=yes, 0=no
CONTRAST1_4	Willing and available to participate, slots open for both Method 1 (BCD) and Method 3 (CD)	1=yes, 0=no
CONTRAST1_5	Willing and available to participate, slots open for both Method 2 (CP) and Method 4 (ODP)	1=yes, 0=no
CONTRAST1_6	Willing and available to participate, slots open for both Method 3 (CD) and Method 4 (ODP)	1=yes, 0=no
PROB_ASSIGN_0	Probability of being assigned to Method 0 (Control)	Any value between 0 and 1
PROB_ASSIGN_1	Probability of being assigned to Method 1 (BCD)	Any value between 0 and 1
PROB_ASSIGN_2	Probability of being assigned to Method 2 (CP)	Any value between 0 and 1
PROB_ASSIGN_3	Probability of being assigned to Method 3 (CD)	Any value between 0 and 1
PROB_ASSIGN_4	Probability of being assigned to Method 4 (ODP)	Any value between 0 and 1

Variables in the randomization dataset can be divided into five major categories (as displayed in Exhibit H8). The definition for each category of variables is described below:

1. PREF0: values of these dummy indicators were obtained from the ‘MTH’ variables and time slots availability indicators in the screening data. If a participant was interested (MTH=1) and available to participate in at least one time slot (at least one slot indicator = 1) of a method, then PREF0=1 for this method.
2. PREF1: values of these dummy indicators were calculated based on PREF0 as well as the number of assigned participants in each deliberative group when a participant was assigned. If PREF0=1 for a method and at least one of the participant’s available groups was open (i.e., had not reached the target number of recruits) within the method, then PREF1=1.
3. CONTRAST0: values of these dummy indicators were calculated based on PREF0. If PREF0=1 for both methods pertinent to a contrast, then the corresponding CONTRAST0=1.
4. CONTRAST1: values of these dummy indicators were calculated based on PREF1. If PREF1=1 for both methods pertinent to a contrast, then the corresponding CONTRAST1=1.
5. PROB_ASSIGN: values of these variables are the probabilities of being assigned to each experimental condition (i.e., four deliberative methods and the control group), thus the values were always between 0 and 1 and the sum of PROB_ASSIGN_0 to PROB_ASSIGN_4 within a participant is always equal to 1.

In summary, variables in the randomization dataset were used to identify each participant’s eligibility for a method or a contrast based on different standards. If different standards were ranked by their levels of strictness from low to high, then we can get:

- $MTH < PREF0 < PREF1$
- $CONTRAST0 < CONTRAST1$

In other words, for a specific method, the participants with PREF1=1 is a subset of those with PREF0=1, and the participants with PREF0=1 is a subset of those with MTH=1. By the same token, for a specific contrast, the participants with CONTRAST1=1 is a subset of those with CONTRAST0=1.

Incorporating Randomization Scheme into Analyses

The key strategy of incorporating randomization scheme into analyses revolved around selecting an appropriate subsample of participants for each analysis. Below, we illustrate how to use the variables in the assignment dataset and the randomization dataset to select subsamples of data for Research Questions 1 and 3, respectively. We used descriptive analyses to answer Research Question 2; therefore, we did not incorporate the randomization scheme into the analyses for that research question.

Research Question 1

Is public deliberation more or less effective than educational materials alone at changing knowledge about the deliberative topic, and is there a concomitant shift in attitudes?

RQ1 investigated the effectiveness of deliberation in changing knowledge and attitudes. It involved comparing all participants who participated in a deliberative method to the control group or comparing participants in all four deliberative methods combined to controls. It estimated the overall effect of deliberation (i.e., the main effect), as well as method-specific effects to estimate the effect of the individual methods.

Sub-questions under RQ1 include:

- RQ1.1. Do participants in all four deliberative methods *combined* demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?
- RQ1.2. Do participants in *each of the four* deliberative methods demonstrate larger changes in knowledge and attitudes about the deliberative topics than the control group members?

For each comparison between deliberative method(s) and the control group, we selected the subsample of participants who were assigned to either of the two pertinent conditions. To be more specific, Exhibit H9 lists the selection criteria for each sub-question of RQ1. All the participants who satisfy the selection criteria of a particular comparison were included in that comparison.

Exhibit H9. Subsample selection criteria for Research Question 1

Research Question	Contrast	Selection Criteria (Based on the variable 'MethodID' in the assignment/attendance dataset)
RQ1.1	All vs. Control	MethodID = 0, 1, 2, 3, or 4
RQ1.2	BCD vs. Control	MethodID = 1 or MethodID = 0
RQ1.2	CP vs. Control	MethodID = 2 or MethodID = 0
RQ1.2	CD vs. Control	MethodID = 3 or MethodID = 0
RQ1.2	ODP vs. Control	MethodID = 4 or MethodID = 0

RQ 1.1 included all the participants in our full sample who completed both the pre-test and post-test of the Knowledge and Attitudes Survey. All the participants assigned to deliberative methods were combined and compared to those assigned to the control group. Similarly, RQ 1.2 analyses included all the participants who were assigned to the control group or assigned to a particular method.

Research Question 3

Are specific deliberative methods more effective than others?

- RQ3.1. Does the intensity of the deliberative method affect knowledge, attitudes, and deliberation quality and experience?
- RQ3.2. Does the mode of deliberation affect knowledge, attitudes, and deliberation quality and experience?

For each comparison between two deliberative methods, we selected the subsample of participants who were willing and available to participate in both pertinent methods. In addition, within each method, at least one of the participant’s available deliberative groups had to be open (i.e., had not reached the target number of participants), so that participants get nonzero probabilities to be assigned to either of the two methods. Exhibit H10 lists the selection criteria for each sub-question of RQ3. All of the participants who satisfy the selection criteria of a particular comparison were included in that comparison.

Exhibit H10. Subsample selection selection criteria for Research Question 2

Research Question	Contrast	Selection Criteria (Based on the variables ‘CONTRAST1’ in the randomization dataset and ‘MethodID’ in the assignment/attendance data set)
RQ2.1	BCD vs. CP	CONTRAST1_1 = 1 and (MethodID = 1 or MethodID = 2)
RQ2.2	CD vs. ODP	CONTRAST1_6 = 1 and (MethodID = 3 or MethodID = 4)

For example, to compare the method with the least intensity (BCD) with the method with greatest intensity (CP), we selected all the participants who were assigned to either BCD (i.e., MethodID=1) or CP (i.e., MethodID=2) with the value of CONTRAST1_1 equal to 1 (i.e., those who were willing and available to participate in both BCD and CP, and their interested slots had not been filled for both methods when they were assigned). This subsample of participants was then used to fit the regression model to examine the difference in effectiveness between BCD and CP.

Weighting Procedure

The randomization process resulted in an approximately equal distribution of participants in each deliberative method; however, it produced a different distribution in the control group on (1) health insurance coverage, (2) income, and (3) racial/ ethnic groups. The difference occurred because we prioritized assigning the priority population members to deliberation early in recruitment—meaning there was a higher proportion of the priority population groups in the deliberation arms than in the control. To adjust for this imbalance, we calculated and applied analytic weights for analyses comparing the deliberative methods to the control group (RQ1). Participants were weighted to the U.S. Census population distribution of the relevant variables in the geographic location from which they were selected. Because RQ3 analyses focus on comparing methods to each other and are restricted to participants who were interested in both pertinent methods in a comparison, the participants used in a comparison were similar to each other, and therefore the weight was not necessary for RQ3 analyses.

We created the weight using the following steps.

Step 1: Identify variables causing imbalance

In this step, we identified all of the variables that were causing the imbalance of assignment/response between treatment group (i.e., deliberative method) and the control group (reading materials only). The following logistic regression model was used:

$$\text{logit}(E) = X_1 + \dots + X_{12}$$

where E is a dummy indicator of the method assignment with 0=control, 1=treatment, and X_1 to X_{12} are demographic variables. X_1 to X_{12} correspond to Q58, Q59, Q60, Q61_rec, Q62, Q63_real, Q65, Q66, Q67, Q68, Q69_4, Q70 in the Knowledge and Attitudes Pre-test Survey data¹.

The following demographic variables were found to be significant predictors of the probability of being assigned to a condition. In other words, these variables were the causes of imbalance across conditions:

- Deliberative Methods (all 4 methods combined) vs. Control:
 - health insurance coverage(Q61_rec), income(Q67), ethnicity(Q68), race(Q69_4)
- Deliberative Method 1 (BCD) vs. Control:
 - health insurance coverage(Q61_rec), education(Q65), ethnicity(Q68)
- Deliberative Method 2 (CP) vs. Control:
 - ethnicity(Q68), race(Q69_4)
- Deliberative Method 3 (CD) vs. Control:
 - income(Q67), ethnicity(Q68), race(Q69_4)
- Deliberative Method 4 (ODP) vs. Control:
 - health insurance coverage(Q61_rec), ethnicity(Q68), race(Q69_4)

Step 2: Define demographic strata

To define demographic strata, we selected the four demographic variables that caused the imbalance between deliberative methods (all four methods combined) and the control group: health insurance coverage (Q61_rec), income(Q67), ethnicity(Q68), and race(Q69_4).

Note that although education was a significant cause of imbalance between the BCD method and the control group, we did not include education when constructing strata, because education was highly correlated with income. Thus, by including income alone, we have already taken into account the influence of education to some degree. In addition, excluding education helps us to reduce the total number of defined strata and avoid the “small cell” problem².

To avoid the “small cell” problem, we also recoded the following demographic variables by collapsing the original response categories into a smaller number of categories:

1. Collapsing race (African American vs. Other) and ethnicity (Hispanic vs. Non-Hispanic) into a single variable with three categories, because there were very few African American Hispanic people in the sample.

¹ Note that Q61_rec is a recoded variable for Q61 (health insurance coverage) where the response categories of ‘No’ and ‘Don’t Know’ are combined as one category, as there were too few participants (n=12) answering ‘Don’t Know’. Q63_real is a recoded variable for Q63 (age), where participants’ real ages were calculated based on the raw survey data (where 18 years old was coded as 1, and 19 as 2, etc.). The variable Q64 (marital status) is excluded to avoid the quasi-complete separation and convergence problem for the logistic regression model.

² Small cells (i.e., strata with only a few participants in them) can be problematic because the bias reduction gets overpowered by an increase in variance.

2. Collapsing the original five categories of income into two larger categories (less than \$60,000 vs. \$60,000 or higher)

In summary, there were a total of 12 strata (2*2*3) with

- health insurance coverage (Q61_rec: yes vs. no/don't know)
- income (Q67_rec: less than \$60,000 vs. \$60,000 or higher)
- racial/ethnic groups (Q6869_rec: Hispanic, African American non-Hispanic and other Non-Hispanic)

Step 3: Calculate population proportion for each stratum (census data)

For each of the three classifying variables used to identify the strata (i.e., Q61_rec, Q67_rec, Q6869_rec), the population proportions from the Census data (averaged across the four geographic locations in this study) are shown in Exhibit H11. Subsequently, for each of the 12 strata, we calculated the corresponding proportion of people based on the Census data, using the following formula. These percentage values were defined as the “population proportion,” *Pr_pop* (see the “Population Proportion” column in Exhibit H11).

$$Pr_pop(Strata_ijk)=Pr(Q61_rec=i)*Pr(Q67_rec=j)*Pr(Q6869_rec=k)$$

For instance, for Strata_111 (i.e., participants who have health insurance coverage, with lower income, and are Hispanic), the population proportion is calculated by 84.95%*51.21%*14.05%=6.11%.

Step 4: Calculate the sample proportion of each stratum (by method) for RQ1

Based on our sample data, the percentage of people in each of the 12 strata, within each of the five experimental conditions (four deliberative methods and one control group), respectively, was then calculated, using the SAS procedure (proc freq). These values were defined as the “sample proportion,” *Pr_samp* (see the ‘Sample Proportion’ columns in Exhibit H11).

Step 5: Calculate post-stratification weight based on population and sample proportions

For each participant, we calculated the post-stratification weight using the formula

$$weight = Pr_pop / Pr_samp$$

The post-stratification weights for the 12 strata within each of the five experimental conditions are listed in the “Weight” columns in Exhibit H11. For instance, if an African American Non-Hispanic who has health insurance coverage and lower income (Strata_112) participated in the CP method, then his/her weight is 0.56. We created a weighting variable named as “*St_byMethodID*” and saved it in the dataset.

Step 6: Use post-stratification weighting variable for outcome analyses

The post-stratification weighting variable was used for two types of analyses:

1. Calculating the weighted mean and standard deviation for each of the knowledge and attitude outcomes within each of five experimental conditions. The ‘proc means’ procedure in SAS was used to obtain the weighted mean and standard deviation. Reporting the weighted mean and standard deviation for each experimental condition allows us to see the “real” pre-post changes in knowledge and attitudes, after correcting the non-response bias (i.e., demographic imbalance between treatment and control), in a weighted sample where the demographic profile is representative of the local population demographics.
2. Estimating the treatment effect (i.e., the effectiveness of deliberation) using weighted regression model. The ‘weight’ statement in the ‘proc surveyreg’ procedure in SAS was used to fit the weighted regression model. For analysis in RQ1, the ‘*St_byMethodID*’ variable (see Step 5) was used. In addition to using post-stratification weighting variable for the regression model, we also included other demographic variables that may influence the outcomes as covariates in the regression model. Even though these characteristics should be balanced across groups, including them in the regression models can further guard against inadequacies in randomization and post-stratification weighting.

Evaluating the Performance of Weighting Procedure

To evaluate the performance of the post-stratification weighting, we compared participants’ weighted baseline survey scores across the five experimental conditions. If the weighting procedure worked well, the participants’ characteristics should be balanced between the treatment groups and the control group, and there should not be any significant difference in baseline knowledge and attitude scores across experimental conditions in the weighted sample. See Exhibit H12 for the weighted results and Exhibit H13 for unweighted results with the *p* values from the comparison of baseline scores between conditions, including a total of five contrasts of research interest. There are a total of 14 outcome variables from the survey, which characterize participants’ health knowledge and attitude in different aspects. Results indicate that the weighting procedure worked well in eliminating the imbalance in baseline scores across experimental conditions. There are fewer bolded values in Exhibit H12, compared to Exhibit H13, which indicates that fewer outcomes showed significant difference in baseline scores between conditions after the weighting approach was used. Exhibit H12 also shows that there is no difference in baseline scores for the majority of outcome variables.

Exhibit H11a. Sample proportion, and population proportion for the 12 strata

Strata #	Category for Classifying Variable Q61_rec	Category for Classifying Variable Q67_rec	Category for Classifying Variable Q6869_rec	Population Proportion for Classifying Variable Q61_rec	Population Proportion for Classifying Variable Q67_rec	Population Proportion for Classifying Variable Q6869_rec	Population Proportion (Pr_Pop)	Sample Proportion (Pr_Samp) Read	Sample Proportion (Pr_Samp) BCD	Sample Proportion (Pr_Samp) CP	Sample Proportion (Pr_Samp) CD	Sample Proportion (Pr_Samp) ODP
111	1	1	1	84.95%	51.21%	14.05%	6.11%	1.15%	2.56%	7.29%	3.42%	4.38%
112	1	1	2	84.95%	51.21%	24.19%	10.52%	10.34%	10.26%	18.75%	14.45%	9.96%
113	1	1	3	84.95%	51.21%	61.76%	26.87%	17.82%	19.05%	14.58%	17.11%	13.94%
121	1	2	1	84.95%	48.79%	14.05%	5.82%	2.59%	4.03%	1.04%	3.42%	3.59%
122	1	2	2	84.95%	48.79%	24.19%	10.03%	6.90%	4.40%	14.58%	7.60%	8.76%
123	1	2	3	84.95%	48.79%	61.76%	25.60%	35.92%	24.18%	17.71%	22.81%	25.50%
211	2	1	1	15.05%	51.21%	14.05%	1.08%	0.86%	2.56%		3.80%	3.19%
212	2	1	2	15.05%	51.21%	24.19%	1.86%	2.59%	10.26%	9.38%	8.37%	6.77%
213	2	1	3	15.05%	51.21%	61.76%	4.76%	9.48%	8.79%	6.25%	6.46%	8.76%
221	2	2	1	15.05%	48.79%	14.05%	1.03%	0.29%	2.56%	2.08%	0.76%	0.80%
222	2	2	2	15.05%	48.79%	24.19%	1.78%	2.01%	1.83%	4.17%	3.04%	1.20%
223	2	2	3	15.05%	48.79%	61.76%	4.54%	10.06%	9.52%	4.17%	8.75%	13.15%

NOTE: Q61_rec: Health insurance coverage, 1=yes, 2=no/don't know

Q67_rec: income, 1=less than \$60,000, 2=\$60,000 or higher

Q6869_rec: ethnicity/race, 1=Hispanic, 2= African American Non-Hispanic, 3=other Non-Hispanic

Exhibit H11b. Sample weight for the 12 strata

Strata #	Category for Classifying Variable Q61_rec	Category for Classifying Variable Q67_rec	Category for Classifying Variable Q6869_rec	Weight (St_byMethodID) Read	Weight (St_byMethodID) BCD	Weight (St_byMethodID) CP	Weight (St_byMethodID) CD	Weight (St_byMethodID) ODP
111	1	1	1	5.32	2.38	0.84	1.79	1.39
112	1	1	2	1.02	1.03	0.56	0.73	1.06
113	1	1	3	1.51	1.41	1.84	1.57	1.93
121	1	2	1	2.25	1.45	5.59	1.70	1.62
122	1	2	2	1.45	2.28	0.69	1.32	1.14
123	1	2	3	0.71	1.06	1.45	1.12	1.00
211	2	1	1	1.26	0.42		0.28	0.34
212	2	1	2	0.72	0.18	0.20	0.22	0.28
213	2	1	3	0.50	0.54	0.76	0.74	0.54
221	2	2	1	3.59	0.40	0.50	1.36	1.29
222	2	2	2	0.88	0.97	0.43	0.58	1.49
223	2	2	3	0.45	0.48	1.09	0.52	0.34

NOTE: Q61_rec: Health insurance coverage, 1=yes, 2=no/don't know

Q67_rec: income, 1=less than \$60,000, 2=\$60,000 or higher

Q6869_rec: ethnicity/race, 1=Hispanic, 2= African American Non-Hispanic, 3=other Non-Hispanic

Exhibit H12. *P* values for comparing weighted baseline scores:
* indicates *p* value < 0.05, statistically significant)

Outcome Variable	RQ 1: Deliberation vs No Deliberation: Control vs. All	RQ 1: Deliberation vs No Deliberation: Control vs. BCD	RQ 1: Deliberation vs No Deliberation: Control vs. CP	RQ 1: Deliberation vs No Deliberation: Control vs. CD	RQ 1: Deliberation vs No Deliberation: Control vs. ODP
Q12	0.306	0.165	0.588	0.444	0.744
Q34	0.275	0.266	0.084	0.122	0.337
Q35	0.788	0.616	0.184	0.733	0.818
Q36	0.108	0.620	0.410	0.192	0.045*
Q37	0.251	0.702	0.228	0.797	0.069
Q38	0.534	0.382	0.391	0.254	0.340
Q39	0.221	0.061	0.622	0.247	0.933
Factor1	0.499	0.925	0.860	0.425	0.365
Factor2	0.353	0.523	0.230	0.166	0.343
Factor3	0.886	0.557	0.853	0.679	0.597
Factor4	0.530	0.851	0.075	0.277	0.282
Factor5	0.578	0.559	0.985	0.841	0.484
Factor6	0.405	0.608	0.357	0.291	0.805
knowledge	0.237	0.873	0.114	0.769	0.076

Exhibit H13. *P* values for comparing unweighted baseline scores:
* indicates *p* value < 0.05, statistically significant)

Outcome Variable	RQ 1: Deliberation vs. No Deliberation: Control vs. All	RQ 1: Deliberation vs. No Deliberation: Control vs. BCD	RQ 1: Deliberation vs. No Deliberation: Control vs. CP	RQ 1: Deliberation vs. No Deliberation: Control vs. CD	RQ 1: Deliberation vs. No Deliberation: Control vs. ODP
Q12	0.133	0.227	0.110	0.180	0.605
Q34	0.257	0.222	0.131	0.127	0.334
Q35	0.183	0.975	0.009*	0.290	0.285
Q36	0.013*	0.256	0.304	0.061	0.002*
Q37	0.121	0.903	0.001*	0.531	0.06
Q38	0.599	0.741	0.448	0.533	0.186
Q39	0.588	0.152	0.846	0.738	0.648
Factor1	0.435	0.180	0.419	0.801	0.576
Factor2	0.628	0.476	0.036*	0.128	0.899
Factor3	0.258	0.440	0.065	0.834	0.330
Factor4	0.874	0.878	0.038*	0.408	0.448
Factor5	0.048*	0.142	0.254	0.188	0.057
Factor6	0.091	0.525	0.034*	0.047*	0.502
knowledge	0.592	0.954	0.391	0.791	0.184

Appendix I. Knowledge and Attitudes Survey

Group Description _____
Group # _____
Location _____

Form Approved
OMB No. 0935-0199
Exp. Date 06/30/2014

COMMUNITY FORUM PRE- AND POST-SURVEY

SURVEY INSTRUCTIONS

Thank you for agreeing to participate in the Community Forum project.

The Agency for Healthcare Research and Quality (AHRQ) is sponsoring the American Institutes for Research to conduct this survey as part of its evaluation of the Community Forum project. It includes questions about your opinions on health care topics and your understanding of health care issues, and some additional questions about you. It will take you about 30 minutes to complete this survey.

- You should fill out this survey only if you have been selected to participate in the Community Forum project.
- Answer all the questions by checking the box to the left of your answer.

Name: _____

ID # _____

Public reporting burden for this collection of information is estimated to average 30 minutes per response, which is the estimated time required to complete the survey. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: AHRQ Reports Clearance Officer; Attention: PRA, Paperwork Reduction Project (0935-0176); AHRQ; 540 Gaither Road, Room # 5036; Rockville, MD 20850.

Participant ID: _____

The first questions ask you for your opinions on how doctors should make treatment recommendations.

- 1.** How important is it that doctors provide treatments that are based on a patient's preferences?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

- 2.** How important is it that doctors provide treatments that are based on their expert medical opinions?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

- 3.** How important is it that doctors provide treatments based on medical research results?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

The next questions ask you for your opinions on factors doctors should consider when making treatment recommendations.

- 4.** How important is it that doctors consider the side effects when recommending treatments?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

- 5.** How important is it that doctors consider how much a person has to pay when recommending treatments?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

- 6.** How important is it that doctors consider how much health insurance companies have to pay when recommending treatments?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

The next questions ask you about your opinions on how results of medical research should be used when you and your doctor make treatment decisions.

7. How important is it for you to know the results of medical research when making treatment decisions with your doctor?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

8. How important is it for you to ask your doctor about medical research results related to your health problem?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

9. How important is it for people to know the results of medical research when making treatment decisions with their doctor?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

10. How important is it that people ask their doctors about the results of medical research related to their health problem?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

11. How important is it that doctors tell you about medical research results that are related to your treatments?

- 1 ☐ Not important at all
- 2 ☐ Not important
- 3 ☐ No opinion
- 4 ☐ Important
- 5 ☐ Very important

12. Some people think the results of medical research should be the most important factor when making a medical treatment decision. Others think doctor's knowledge about the patient is the most important factor. Some people think both factors are important but one should get more weight than the other. On a scale from 1 to 5, where 1 means you believe that only medical research counts and 5 means that only the doctor's knowledge counts, mark the number that indicates where you fall on this scale.

Results of medical research	1	2	3	4	5	Doctor's knowledge about me
--------------------------------------	---	---	---	---	---	-----------------------------------

13. In your opinion, which of the following best determines when medical treatments are "proven to be effective":

- 1 ☐ "Proven to be effective" is best determined by the results of medical research.
- 2 ☐ "Proven to be effective" is best determined by each doctor based on his or her own experience.
- 3 ☐ "Proven to be effective" is best determined by people who know their bodies best and can say what is effective and what is not.
- 4 ☐ No opinion

In the next section, please indicate whether you agree with the following statements about the use of medical research results when making treatment decisions.

14. Doctors should always make treatment recommendations based on medical research results, even though research is always being updated.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

15. Doctors should be able to provide a medical treatment even if medical research results show it won't work for the patient.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

16. Doctors should be able to provide a medical treatment even if medical research results show it might not work for the patient.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

17. Doctors should be able to provide a medical treatment they prefer, even if medical research shows that there is another treatment that works better.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

18. People should be able to get a medical treatment they prefer, even if medical research results show it won't work for them.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

19. People should be able to get a medical treatment they prefer, even if medical research results show it might not work for them.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

20. People should be able to get a medical treatment they prefer, even if medical research shows there is another treatment that works better.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

21. People should be able to get any medical treatment, no matter how much the treatment costs.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

22. Doctors should be able to provide any medical treatment, no matter how much the treatment costs.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

23. Doctors should be able to provide an expensive medical treatment, even if medical research shows there is a treatment that costs less and works the same.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

24. For health problems that are not life threatening, doctors should be able to recommend any medical treatment, no matter how much the treatment costs.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

25. Health insurance should pay for a treatment that a person prefers, even if medical research results show that it does not work.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

26. People with health insurance should consider the effect of their treatment decisions on the cost of health insurance premiums for everyone in their health plan.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

The next questions ask you for your opinions on whether parents, schools, or the government should limit the consumption of high-fat, high calorie foods.

Medical research has shown that many Americans consume too much unhealthy food that is high in fat and high in calories. Medical research results indicate that these behaviors can lead to obesity and further health problems, such as heart disease and diabetes.

How much do you agree or disagree with the following statements:

27. Parents should limit how much unhealthy food their children eat.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

28. Schools should limit how much unhealthy food is served to children.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

29. Teachers should teach their students healthy eating habits.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

30. The government should limit advertising of unhealthy food to children.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

31. The government should tax unhealthy food to discourage people from buying it.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

32. The government should limit the sale of unhealthy foods in the United States.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

33. The government should ban the sale of food that research has shown to cause heart disease.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

The next questions ask you for your opinions about the use of antibiotics, a type of medicine used to treat some types of infections.

Antibiotics treat infections caused by bacteria. But, antibiotics do not work against infections caused by viruses. Sometimes it is difficult to tell whether the infection is viral or bacterial without a special test. When antibiotics are used too often or for the wrong reasons, bacteria become stronger and the antibiotics stop working. This means antibiotics may not work in the future for people with infections caused by bacteria.

How much do you agree or disagree with the following statements:

34. People should be able to get an antibiotic if they want it, even if it might not help.

1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

35. The doctor should be able to prescribe an antibiotic to treat infections, even if there is a small chance it would help.

1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

36. The government should limit when doctors can prescribe antibiotics.

1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

The next set of questions asks you for your opinions about how successful hospitals are at performing surgeries.

Medical research results show that success rates for some types of surgeries are better in hospitals that do a lot of surgeries (called high-volume hospitals) compared to hospitals that do fewer surgeries (called low-volume hospitals). For example, 4 out of 100 children die from heart surgery in high-volume hospitals. In contrast, 15 out of 100 children die from heart surgery in low-volume hospitals.

How much do you agree or disagree with the following statements:

37. Insurance companies should require children who need heart surgery to be treated at high-volume hospitals.

1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

38. Hospitals should be required to achieve a high success rate in order to continue performing heart surgery on children.

1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

39. Doctors who treat patients at low-volume hospitals should tell their patients if other high-volume hospitals have better results.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Neither disagree nor agree
- 4 ☐ Agree
- 5 ☐ Agree strongly

Please indicate how much you oppose or support the following statements:

40. People who are obese have more health problems and live shorter lives than people with healthy weights. How much do you support the government spending money to help obese people lose weight?

- 1 ☐ Strongly oppose
- 2 ☐ Oppose
- 3 ☐ Neither oppose nor support
- 4 ☐ Support
- 5 ☐ Strongly support

41. How much do you support the government spending money to help people stop smoking?

- 1 ☐ Strongly oppose
- 2 ☐ Oppose
- 3 ☐ Neither oppose nor support
- 4 ☐ Support
- 5 ☐ Strongly support

The purpose of the next set of questions is to help us understand what people know about medical research. *You may not know all the answers. That is okay.* We are just trying to learn more about how much people understand about medical research:

42. For a new medicine to be approved for use in the United States, medical research results have to show that:

[MARK ALL THAT YOU THINK ARE TRUE.]?

- 1 ☐ The new medicine works better than medicines already approved for use
- 2 ☐ The new medicine is effective
- 3 ☐ The new medicine is safe
- 4 ☐ The new medicine is approved in other countries
- 5 ☐ Don't know

43. Comparative Effectiveness Research is a type of research that can:

[MARK ALL THAT YOU THINK ARE TRUE.]?

- 1 ☐ Tell us which existing treatments work better
- 2 ☐ Tell us which existing treatments have fewer side effects
- 3 ☐ Discover new treatments
- 4 ☐ Test new treatments
- 5 ☐ Don't know

44. Who pays the largest portion of total health care costs in the United States?

- 1 ☐ Individuals and families
- 2 ☐ Government (Federal, State, and local)
- 3 ☐ Private businesses, including health insurance companies
- 4 ☐ Other (e.g., charities)
- 5 ☐ Don't know

45. The results of Comparative Effectiveness Research can be used to develop quality standards for treating people.

- 1 ☐ Definitely true
- 2 ☐ Probably true
- 3 ☐ Probably false
- 4 ☐ Definitely false
- 5 ☐ Don't know

46. When people in health plans get more care than they need (called unnecessary care), this affects the overall costs for everyone in the health plan.

- 1 ☐ Definitely true
- 2 ☐ Probably true
- 3 ☐ Probably false
- 4 ☐ Definitely false
- 5 ☐ Don't know

47. More expensive medical treatments work better than less expensive treatments.

- 1 ☐ Definitely true
- 2 ☐ Probably true
- 3 ☐ Probably false
- 4 ☐ Definitely false
- 5 ☐ Don't know

48. Doctors usually make treatment recommendations based on medical research results.

- 1 ☐ Definitely true
- 2 ☐ Probably true
- 3 ☐ Probably false
- 4 ☐ Definitely false
- 5 ☐ Don't know

49. Hospitals that perform many hip replacement surgeries each year get better results than hospitals that perform fewer hip surgeries.

- 1 ☐ Definitely true
- 2 ☐ Probably true
- 3 ☐ Probably false
- 4 ☐ Definitely false
- 5 ☐ Don't know

In the next section, please indicate your opinion on health care costs for each question.

50. Which of the following statements comes closer to your own view?

- 1 ☐ The government or health insurance plans should pay for any drug or medical treatment that has been approved as being safe and effective for saving lives or improving people's health, regardless of what is costs.
OR:
- 2 ☐ There are so many new, expensive prescription drugs and medical or surgical treatments that it is too expensive for government or health insurance plans to pay for all of them.

51. In many countries there are concerns about slowing the growth of health care costs. It has been suggested that some medical or surgical treatments that have been shown to be safe and effective should **not** be paid for by the government or health insurance plans because their high cost is not felt to be justified by the amount of benefit they provide. Do you favor or oppose doing this?

- 1 ☐ Favor
- 2 ☐ Oppose

52. Do you favor or oppose the U.S. having a government decision-making body that recommends whether government programs should pay for medical or surgical treatments if they think they cost too much?

- 1 ☐ Favor
2 ☐ Oppose

53. Do you favor or oppose the government or your health insurance plan paying for a more expensive medical or surgical treatment recommended by your doctor even if it has not been shown to work better than less expensive treatments?

- 1 ☐ Favor
2 ☐ Oppose

54. In another country, the national government decided against paying for a new drug for treating an advanced form of cancer. On average, the drug costs \$35,000 per patient. The drug does not cure the disease, but studies suggest that using the drug can add, on average, about six months to a patient's life. Some patients would gain only a short period, while others could gain a lot more time. If this decision **not** to pay for this drug were made in the U.S., would you approve or disapprove of the decision?

- 1 ☐ Approve
2 ☐ Disapprove

The next set of questions asks how much you agree or disagree with statements about your health and healthcare.

55. When all is said and done, I am the person who is responsible for managing my health.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

56. Taking an active role in my own health care is the most important factor in determining my health and ability to function.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

57. I am confident that I can take actions that will help prevent or minimize health problems.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Neither disagree nor agree
4 ☐ Agree
5 ☐ Agree strongly

The last set of questions asks your age, race, ethnicity and some additional questions. These questions help us make sure that people with different backgrounds get included in the project and help us understand how different people think and feel about the healthcare issues discussed in the Community Forum Project.

58. In general, how would you rate your overall health?

- 1 ☐ Excellent
- 2 ☐ Very good
- 3 ☐ Good
- 4 ☐ Fair
- 5 ☐ Poor

59. In the past 12 months, have you seen a doctor or other health care provider 3 or more times for the same condition or problem? Do not include pregnancy.

- 1 ☐ Yes
- 2 ☐ No

60. In the past 12 months, has a close friend or family member seen a doctor or other health care provider 3 or more times for the same condition or problem? Do not include pregnancy.

- 1 ☐ Yes
- 2 ☐ No

61. Do you have any kind of health care coverage, including health insurance, prepaid plans (such as HMOs), government plans (such as Medicaid or Medicare), or health care coverage through the Veterans Administration or military.

- 1 ☐ Yes
- 2 ☐ No
- 3 ☐ Don't know

62. Are you male or female?

- 1 ☐ Male
- 2 ☐ Female

63. How old are you?

Age in years _____

64. What is your current marital status?

- 1 ☐ Never married
- 2 ☐ Living with a partner
- 3 ☐ Married
- 4 ☐ Widowed
- 5 ☐ Divorced
- 6 ☐ Separated

65. What is the highest grade or level of school that you have completed?

- 1 ☐ 8th grade or less
- 2 ☐ Some high school, but did not graduate
- 3 ☐ High school graduate or GED
- 4 ☐ Some college or 2-year degree
- 5 ☐ 4-year college degree
- 6 ☐ More than 4-year college degree

66. How would you describe your employment status at this time?

- 1 ☐ I am employed for pay full-time or part-time.
- 2 ☐ I am currently unemployed or employed less than full-time and I am looking for work.
- 3 ☐ I am not employed for pay but I am a student, a homemaker, or volunteer.
- 4 ☐ I am retired.
- 5 ☐ My health does not permit me to be employed for pay.

67. Using the categories below, please indicate the annual income of your household (include yourself and anyone with whom you live and have common finances):

- 1 ☐ Less than \$15,000
- 2 ☐ \$15,000 to \$29,999
- 3 ☐ \$30,000 to \$59,999
- 4 ☐ \$60,000 to \$100,000
- 5 ☐ More than \$100,000

68. Are you Hispanic or Latino/Latina?

- 1 ☐ Yes
- 2 ☐ No

69. What is your race/ethnicity? Please select one or more.

- 1 ☐ American Indian or Alaska Native
- 2 ☐ Asian
- 3 ☐ Native Hawaiian or other Pacific Islander
- 4 ☐ Black or African American
- 5 ☐ White

70. Do you speak a language other than English at home?

- 1 ☐ Yes
- 2 ☐ No

71. (If yes) What is this language?

72. (If yes) How well do you speak English?

- 1 ☐ Very well
- 2 ☐ Well
- 3 ☐ Not well
- 4 ☐ Not at all

Appendix J. Deliberation Quality and Experience Survey

Group Description _____
Group # _____
Location _____

Form Approved
OMB No. 0935-0199
Exp. Date 06/30/2014

YOUR COMMUNITY FORUM EXPERIENCE

SURVEY INSTRUCTIONS

- You should fill out this survey only if you have been selected to participate in the Community Forum project. Do not fill out this survey if you will not participate in this project.
- Answer all the questions by checking the box to the left of your answer.

Place removable label here.

Public reporting burden for this collection of information is estimated to average 15 minutes per response, which is the estimated time required to complete the survey. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: AHRQ Reports Clearance Officer; Attention: PRA, Paperwork Reduction Project (0935-0176); AHRQ; 540 Gaither Road, Room # 5036; Rockville, MD 20850.

Participant ID: _____

Please answer the questions below about your experience in taking part in the Community Forum project discussion group. If your group met more than one time, please answer the questions about your overall experience.

1. People in the group respected my opinion.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

2. People in the group really listened to each other.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

3. Even when people disagreed, they respected each other's opinions.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

4. There was too much disagreement in my group.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

5. Some people in the group spoke a lot more than others.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

6. Some people in the group spoke too much.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

7. Some people in the group barely spoke at all.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

8. I spoke as much as I wanted to in the group.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

9. By the end of the discussion, I understood why people held certain beliefs that were different from mine.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

10. The discussion led me to change some of my opinions.

- 1 ☐ Disagree strongly
2 ☐ Disagree
3 ☐ Agree
4 ☐ Agree strongly

11. People gave reasons to support their opinions.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

12. Attending the discussion today was good use of my time.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

13. I would participate in activities like this in the future.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

14. More events like this should be held as a way of getting the views of people throughout the country.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

15. How much did the group discussions affect your opinions on the issues?

- 1 ☐ Not at all
- 2 ☐ A little bit
- 3 ☐ A lot

16. How much did the educational materials affect your opinions on the issues?

- 1 ☐ Not at all
- 2 ☐ A little bit
- 3 ☐ A lot

17. How much did the sessions with the experts affect your opinions on the issues?

- 1 ☐ Not at all
- 2 ☐ A little bit
- 3 ☐ A lot
- 4 ☐ There were no experts

18. How did you feel during the group discussions?

- 1 ☐ Not at all interested
- 2 ☐ Not very interested
- 3 ☐ Somewhat interested
- 4 ☐ Very interested

19. How did you feel during the presentations?

- 1 ☐ Not at all interested
- 2 ☐ Not very interested
- 3 ☐ Somewhat interested
- 4 ☐ Very interested

20. I think the recommendations made by the group will influence how research findings are used.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

21. The event was well organized.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

22. There was enough time to fully discuss all the relevant issues.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

23. I understood what I was asked to do throughout the discussion group.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

24. The facilitator sometimes tried to influence the group with his or her own opinions.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

25. I could tell that the facilitator favored some positions over others.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

26. The facilitator made sure that all opinions were considered.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

27. The facilitator encouraged everyone to have his or her say.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

28. The information presented was clear and easy to understand.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

29. There was so much information it was difficult to absorb it all.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

30. I understood the issues well enough to participate as much as I wanted to.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

31. The materials presented all sides of the issues.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

32. The experts provided credible information.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly
- 5 ☐ There were no experts

33. The purpose of the event was clear.

- 1 ☐ Disagree strongly
- 2 ☐ Disagree
- 3 ☐ Agree
- 4 ☐ Agree strongly

Appendix K. Summary of Responses to Case Studies

To help participants tackle complex topics, we developed specific examples or “case studies” to illustrate the issues and promote deliberation. We used five case studies—*Comparing Hospital Quality (Hospital Quality)*, *Upper Respiratory Infections in Children: Antibiotics Versus Symptom Treatment (URI)*, *Obesity Management: Comparing Prevention and Treatment (Obesity Management)*, *Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicines (Heart Disease)*— and a hypothetical example, *Comparing Approaches to Preventing Illness: A Fictional Case (Fictional Case)*. This appendix provides background information for, and a summary of participant responses to, each case study. Appendix C has the full case studies.

Case Study 1. Comparing Hospital Quality

This case study presented information on the differences in clinical results between high-volume and low-volume hospitals. All methods used this case study to discuss the overall deliberative question. Brief Citizens’ Deliberation (BCD) and Online Deliberative Polling (ODP) discussed only this case study. Community Deliberation (CD) discussed this case study in Session 2 after discussing the URI case study in Session 1. Citizens’ Panel (CP) discussed this case study toward the end of the second day of deliberation.

BCD and CD methods presented the case study in two separate parts. The first part examined using evidence of clinical results in determining whether individuals should get care at a low- or high-volume hospital. Participants responded to the following prompt statements:

Given the differences in results for some patients in these low- and high-volume hospitals, which of the three statements is closest to your view?

- How hospital staff treat a person (such as being respectful, communicating well, etc.) is more important than the differences in clinical results.
- Regardless of what the evidence shows, people should trust that their doctors will provide the best care no matter which hospital they go to.
- People should use only the hospitals that get the best clinical results.

The second part of the case study asked participants to assume the role of societal decisionmakers as members of a town council who had to choose between two health plans:

Health Plan A (low-volume hospital). This plan covers all the services available at Springview Community Hospital, the low-volume, local hospital. The only time services are covered for the high-volume Regional Medical Center 50 miles away is for specialized care that is not available at Springview Community Hospital.

Health Plan B (high-volume hospital). This plan covers all services at the larger Regional Medical Center 50 miles away, which has better clinical results for several specialized surgeries. This plan pays for services at Springview Community Hospital for two situations: (1) maternity care and (2) emergency care before transferring patients to Regional Medical Center.

In the case study, the town council must choose one of these health plans for local government employees. As the local government is the town’s largest employer, choosing Plan B would

likely cause the local, low-volume hospital to lose many patients and eventually close. Given this information, participants deliberated about which plan to choose.

Of note, the first four groups of BCD in Chicago also received a section on costs between these two parts. However, facilitators removed this discussion from later groups due to time limitations and because cost considerations often arose spontaneously.

The **ODP method** presented the case study in three parts. As with BCD and CD, the first part examined using evidence of clinical results in determining whether individuals should get care at a low- or high-volume hospital. The second part assessed cost considerations with low- and high-volume hospitals. The third part asked participants to assume the role of societal decisionmakers as members of a town council who had to choose between two health plans. The materials for the ODP groups did not include the prompt statements. Instead, the case study included pro/con tables and a list of discussion questions after each part of the case study.

The **CP method** included the same information as BCD and CD, but groups discussed the case study as a whole (not broken into parts). Also, this method did not use the facilitator prompts after the first part of the case study.

Main Themes That Emerged During Deliberation

The following main themes emerged from deliberation of this case study:

The *Hospital Quality* case study was extremely challenging for many participants, and some misunderstood key points that could have affected their responses. Most participants understood the overall concept of high- and low-volume hospitals; however, several participants struggled with key aspects of the case study. For example, all participants may not have understood some of the nuances between high- and low-volume hospitals (e.g., a hospital might be high-volume for one procedure but low-volume for another). Also, a common misunderstanding was that Plan B (regional, high-volume hospital) would require participants to travel 50 miles for emergency care. Additionally, some participants appeared to confuse common ailments that might require a hospital visit (e.g., pneumonia) with ones that would be treated in an outpatient, primary care facility (e.g., a sinus infection). As a result, many of these participants chose Plan A (local, low-volume hospital) because they found it unacceptable to travel 50 miles for emergency or primary care.

Several participants appeared to believe (but did not explicitly express) that high-quality care also comes at the highest cost. For example, some participants who indicated that quality was most important assumed that it would consequently come at the highest expense. Many participants who emphasized cost in their decisionmaking also hinted, but did not explicitly articulate, that saving money would come at the expense of quality. For example, one participant noted that *“If we look at cost, it might be a hospital that’s cutting costs, but the care might not be that great,”* (ODP group, remote). In general, some participants appeared to struggle with the concept that the highest-quality care can also be less expensive than other options.

While discussing the last part of the hospital quality case study (choosing between Plans A and B), many participants also expressed the fundamental belief that insurance plans with more provider choices were better. However, participants were divided on whether or not they would be willing to pay more for everyone to have more choices. Some participants were more comfortable with, but not necessarily more supportive of, payers setting boundaries around in-

network providers as long as patients could pay the cost difference for out-of-network care. Some participants would also be willing to pay for a better insurance plan that has more comprehensive coverage. When prompted by the facilitator to address the scenario in which everyone's premiums rise to accommodate more hospital choices, some participants were comfortable with an increase of up to \$25. Other participants did not want their premiums to rise, particularly if it allowed others to make "poor" choices such as selecting less effective care or making poor lifestyle choices. Similarly, many participants noted that patients who disregard evidence during medical decisionmaking should incur the cost of doing so.

Participants discussed factors that would be important when choosing between a high- and low-volume hospital; these comments fell into two groups: (1) selecting a provider based on evidence of clinical results and (2) selecting a provider based on complex considerations in which evidence of clinical effectiveness is not always the most important factor.

1. ***Selecting a provider based on evidence of clinical results.*** Many participants claimed that evidence is the most important factor influencing their personal medical choices, especially when discussing the first part of the case study (using evidence of clinical results in determining whether individuals should get care at a low- or high-volume hospital). Most of these participants were adamant that they make decisions based on the best health care. As one participant explained, "*I mean, I'm not ordering a cheeseburger [where] I don't care what the reviews on Yelp says, I'm going to eat it anyway. But this is surgery, I gotta go to the best place,*" (CD group in Chicago, IL). Participants also explained that that they would do whatever was necessary to pay for or access the best care possible.

When evaluating the three prompt statements presented in Part 1 (BCD and CD groups only), many participants expressed that although personal comfort and convenience are important, clinical results and medical effectiveness trump these considerations when choosing care. Participants in the CD method often emphasized using evidence in decisionmaking, perhaps because they deliberated on this case study after spending Session 1 discussing the impact of antibiotic overuse when evidence is not always used in patient/doctor decisionmaking. Further, when weighing better clinical results against respectful and communicative hospital staff, several participants noted that clinical results matter more than how staff treat patients for serious conditions or surgeries.

2. ***Selecting a provider based on complex considerations in which evidence of clinical effectiveness is not always the most important factor.*** These participants agreed that quality of care is always important in medical decisionmaking, but they acknowledged other factors that are also important, presented below.

Distance. In the case study, the high volume hospital is 50 miles away, and participants often referred to the low-volume hospital as the more convenient option in terms of location. Many participants were sensitive to the logistical and financial constraints, for individuals and their families, resulting from travelling long distances to attain care.

Weighing cost, risks, and benefits. Some participants weighed personal costs, risks, and benefits associated with choosing between a high- and low-volume hospital, indicating they would pick "*the best that I can afford*" (CD group in Durham, NC). For example, one participant explained that "*if the cost is 50 percent more to go to a high-volume [hospital], then I might have to be willing to weigh that risk and go to the low-volume*

hospital” (ODP group, remote). Participants noted that for minor health care issues they might choose a low-cost provider regardless of slightly poorer quality. For major surgeries or treatments, these participants would be willing to pay more for high-volume hospitals to obtain higher quality. Additionally, some participants also mentioned that the high-volume hospital had better clinical outcomes only for certain, specialized procedures. Therefore, these participants would be willing to consider the low-volume hospital for treatment.

Some participants also noted that the immediate, out-of-pocket travel expenses (e.g., gas) were their primary reasons for choosing Plan A, where they have access to local care. Some participants also weighed travel costs for their family. As a result, participants considered the severity of their condition (a complicated surgery versus more routine check-ups) when deciding if the potential benefit from going to a high-volume hospital would be worth the cost and/or travel.

Individual doctor’s performance. Some participants stated that an individual doctor’s success rates determine the quality of care, rather than whether the hospital is low- or high-volume. For example, one participant explained *“It’s not a hospital that gets the clinical results, it’s the doctor that gets the clinical result. The hospital [is] just a building,”* (CD group in Durham, NC).

Support from friends and family. When deciding between the low- and high-volume hospital from an individual perspective, a few participants were willing to choose the low-volume hospital (with less successful clinical outcomes) if it meant that they would be closer to friends and family. As mentioned in the case study, friends and family often provide emotional support when patients are ill.

Patient-centeredness. Participants noted how a patient’s comfort during recovery and time in a hospital affect treatment outcomes. Additionally, some participants worried that they would be just a “number” at a larger, high-volume hospital and would not be given proper attention. Participants described respect and care from hospital staff with the following phrases: “customer service,” “good communication,” “consumer happiness,” “treat me with respect and courtesy and keep me up to date,” and “responsive to your needs.”

Participants believed limited or no information was available to them about provider quality and noted this information would be valuable to make informed decisions. Although participants discussed patients’ responsibility to get information about evidence for certain medical treatments (mostly in the *URI* case study) there was not as much discussion about who should be responsible for seeking statistical information on provider quality. Instead, many participants discussed seeking advice from family or friends when choosing a provider. Some participants also discussed the need for transparent information on provider quality as they believed it was not available to patients. These participants explained that *“the information must be provided to each and every patient so they can make an informed choice,”* (ODP group, remote). Another participant elaborated saying, *“I think it’s important to be provided with information on...what your facility’s track record is. So say if I’m going in and I’m having bypass surgery, I want to know what my hospital’s track record is on that versus a neighboring hospital or other local hospitals,”* (BCD group in Sacramento, CA). In the ODP groups particularly, participants emphasized that the government should require hospitals to report cost

and quality information (e.g., via a transparent Web site). Some participants preferred having this transparent cost and quality information instead of financial incentives to choose cheaper physicians. Further, although a few participants noted that providers should share information on their quality, others responded saying it was unlikely that providers would divulge information if their practice was not the highest in quality.

Participants did not want boundaries to limit provider choice. Participants typically responded to the term “boundaries” as prohibitive in nature, meaning they would not be able to make their own choices. Therefore, many participants wanted to personally choose the best provider for their family and did not want boundaries to limit that decision. Different from the other deliberative methods, the ODP method did not discuss the overarching question, which prompted about boundaries, in the initial session. The ODP facilitator’s guide asked whether providers should meet certain “standards” in one of their discussion questions for Session 1; participants often responded that hospitals should meet certain minimum requirements to continue operating.

When discussing the plans available to the town council, participants often prioritized the economic impact on the community over the evidence of clinical results. Deliberating from a societal perspective, many participants struggled to weigh evidence against who else, besides themselves, their decisions would impact. Several participants noted that they would choose differently, between Plans A and B, based on whether they were picking for themselves versus picking as a town council member. Many participants said they would choose Plan B (high-volume hospital) for themselves and Plan A (low-volume hospital) for the community. Although these participants valued high-quality care for themselves, they chose the low-volume hospital because closing it would affect the entire community. Some participants noted that the economic impact and loss of local access to care was sufficient to outweigh clinical quality. One participant explained that *“If they send everybody to this high-volume hospital and they lay off half the workforce of their town, that’s going to have repercussions on their bottom line....I think they would have to think about how this would financially impact the town as a whole,”* (ODP group, remote). Additionally, many participants expressed that other community members would prioritize economic impact and would consequently not re-elect town council members who voted for Plan B. For example, one participant explained, *“I would have... [picked Plan] B except you made me town council and I had to go with making the money and getting re-elected,”* (CD group in Sacramento, CA). In contrast, a few participants prioritized giving every member of the town the highest chance of living a long, healthy life as more important than job loss in the community; these participants argued that getting the best clinical care is in society’s best interest.

Participants also had other concerns with selecting the high-volume hospital plan:

Concerns about physical access to care. Several participants were also attuned to the needs of those in more rural areas who do not have as much physical access to health care facilities. These participants struggled with imposing restrictions that might affect access to care for others.

Wanting to improve quality at the low-volume hospital. In many groups, a few participants resisted having to make a choice between the options offered and preferred an option aimed to improve quality at the low-volume hospital. These participants generally wanted to create

a “Plan C,” where patients get the highest-quality care at the local hospital to prevent it from closing.

Some participants believed that third-party payers have the “right” to set boundaries on provider choice; however, participants also believed that these payers could put profit ahead of quality when setting boundaries. Discussing the hospital quality case study unearthed many participant beliefs and attitudes about third-party payers’ setting boundaries around provider choice, which we explain below.

Payers have a right to set boundaries. Many participants who acknowledged that payers often absorb the majority of health care costs explained that payers have the right to set some boundaries. Similarly, a minority of participants stated that employers have the right to make choices for their employees, citing four main reasons: (1) employers are invested in their employees and have their best interests in mind; (2) employers pay for the insurance; (3) employers are running a business and have a right to consider the bottom line; and (4) an employee chooses where to work and thus also chooses their level of health insurance benefit. Importantly, there were participants who noted they wanted to retain freedom of choice no matter who pays the majority of their health care costs.

Discomfort with payers’ setting boundaries due to profit motivations. Despite this “right,” many participants said they were uncomfortable with payers’ setting boundaries because profit motivations could lead to decisions that put profit or cost-cutting ahead of quality care. For instance, participants noted that payers might set boundaries on provider choice based on cheapest costs instead of on quality care. Also, employers might decide to offer health insurance through companies with the cheapest plans instead of ones that offer high-quality services. Also, participants said it was unlikely that health care providers would refer patients to another hospital for services they offer, even if the quality might be better, because that is like a car salesman “*sending somebody from a Chevrolet dealership to a Ford dealership*” (ODP group, remote). For these reasons, participants noted that the authorities who set boundaries on health care services should not have a financial interest in the outcome of decisions; some participants commented that medical associations might be best to set these boundaries.

Participants said they were more comfortable with payers setting boundaries if the goals were to lower costs and improve health care quality. Some participants said they were more comfortable with payers’ setting boundaries if their decisions corresponded with ones they would make. Of note, participants did not express how they would be reassured of payers’ or providers’ true motives for setting boundaries that reduce costs. For example, one participant noted: “*If I could be guaranteed that they were picking the best choice, then I would have less of a problem with it. But I’m not always convinced that they’re looking out for the best result.... I believe that [they] are...always looking out for the bottom dollar,*” (BCD group in Silver Spring, MD).

Case Study 2. Upper Respiratory Infections (URI) in Children: Antibiotics Versus Symptom Treatment

We used this case study in two methods: Community Deliberation (during Session 1) methods and the Citizens' Panel (first case study during Day 1).

Community Deliberation (CD) method. The *URI* case study had two parts, and participants in Community Deliberation discussed these two parts separately during the first session. Participants continued the *URI* discussion online between their two sessions, and re-addressed the issues at the beginning of the second session. In the first part, participants weighed the medical evidence of antibiotic overuse in children for *URI* with parent and physician autonomy to use antibiotics. Facilitators asked participants to choose from the following statements after reading Part 1:

- If the doctor thinks that an antibiotic will help the child, then that is more important than what the evidence says.
- It is the parents' decision about getting the antibiotic or not; they care the most for their children and they should decide if they think the benefit of the antibiotic is worth the risk or not.
- Since we know that too many antibiotics might harm the child in the long run, doctors should not prescribe antibiotics unless they know for sure the child has a bacterial infection.

In the second part of the *URI* case study, participants learned about the Methicillin-resistant *Staphylococcus aureus* (MRSA) infection as a societal impact of overusing antibiotics. In this part of the case study, facilitators examined when participants prioritized societal interests over individual doctor and/or patient beliefs, and if these societal interests required definitive actions. Facilitators asked participants to choose from the following statements after reading Part 2:

- I think it is up to the doctor and parents to decide if the antibiotic should be used in each individual case. Parents should not feel responsible for what might or might not happen in the future for the community as a whole.
- I think that given the risks to many people because of overuse of antibiotics, doctors and patients should consider the best interests of the community over individual interests and keep the use of antibiotics down.
- To avoid this problem, there should be stricter rules for when a doctor can order an antibiotic for a patient. We cannot always depend on people to do "the right thing" voluntarily.

Citizens' Panel (CP) method. Participants in Citizens' Panel discussed both parts of the *URI* case study together, focusing on whether the evidence of harm from overusing antibiotics warranted societal boundaries. After presenting the case study, facilitators asked participants about their reactions (i.e., the participants did not receive the prompt statements above). As needed, facilitators also asked about problems with overuse of antibiotics, whether society should set boundaries, and if so, who was responsible for setting those boundaries. In Citizens' Panel, participants broke into groups to discuss the *URI* case; however, all breakouts occurred in the same large room. As a result, many of these discussions could not be captured in transcripts due to background noise.

Main Themes That Emerged During Deliberation

The following main themes emerged from deliberation of this case study:

In the discussion following Part 1 of the *URI* case study, most participants acknowledged that overusing antibiotics is a problem, but some participants questioned whether harm would result from using antibiotics. Also, a few participants argued that their knowledge of their own bodies and their children's personal histories may sometimes override evidence of harm from overusing antibiotics. One participant noted, *"If my child's sick and the doctor is prescribing an antibiotic and I know that antibiotic is going to help my child, I'm not too concerned about what may or may not happen ten years from now. I'm only concerned about getting my child well,"* (CD group in Silver Spring, MD).

At the beginning of the Part 1 discussion in CD groups and in the overall CP group discussions, many participants supported education to increase awareness of the effects of overusing antibiotics to ultimately change patient and provider behavior. Several participants were unaware of the problem before reading the case study; for example, one participant said, *"I think it's a lack of education on my part. I didn't know. I had no clue it was dangerous,"* (CD group in Sacramento, CA). These participants identified the need for increased awareness and education about antibiotic overuse. However, participants differed on how to educate patients. A slight majority of participants suggested that the doctor should explain why the child should not have antibiotics. Another slight minority of participants suggested a wider, public service announcement-style campaign to educate both providers and patients. More participants supported a wider educational campaign if it showcased that physicians have not been following guidelines. Similarly, participants did not perceive patients as responsible for knowing the consequences of antibiotic overuse. But, participants noted that patients should be responsible for asking the provider questions to make an informed decision. Overall, these participants assumed that parents who were aware of MRSA risks would not push for unnecessary antibiotics, and providers who were aware of MRSA risks would follow the recommended care. Participants did not spontaneously mention shortcomings with education, even though the case study mentioned that the American Academy of Pediatrics previously conducted antibiotic use educational campaigns for physicians.

Participants often struggled between protecting the child and protecting the community.

This tension was particularly clear when facilitators presented Part 2 of the *URI* case study in CD groups and the discussion shifted in perspective from the individual to society. This delineation existed, but was less clear, in the CP method as groups discussed both parts at once.

*Protecting the child from harm in Part 1 of the *URI* case.* Participants frequently discussed wanting their child to feel better immediately and were especially concerned about the harms arising from not being able to receive necessary antibiotics. Participants responded more viscerally to the prospect of harm to children and were admittedly more willing to make emotional decisions. Also, participants stated that individuals should have access to the health care that is best suited to their situation, even if it potentially exacerbates a communal burden.

Protecting the community from harm in Part 2 of the URI case. Participants appeared willing to compromise patient/doctor autonomy in certain circumstances and weighed the impact of antibiotic overuse on others and on the future efficacy of antibiotics. In the face of a serious public health issue, society may have the authority to set boundaries. In this context, participants also debated whether or not the overuse of antibiotics specifically was a public health issue.

Attitudes toward regulations in this case study typically fell within three groups over both parts of the case study in Community Deliberation sessions: 1) participants who were initially comfortable with regulation; 2) participants who became more accepting of regulation as deliberation progressed; and 3) participants who firmly did not want regulation. Within each of these groups, participants frequently cited the feasibility of regulations to justify less or no regulation.

1. Participants who were initially comfortable with regulation frequently drew analogies to public health issues that have been resolved through regulation but only after many people were harmed (e.g., narcotic prescription control). Most of these participants did not argue that regulation is inefficient or unfeasible as they cited what they considered successful examples of regulation. Often, these participants said people cannot be trusted to do the “right thing.”
2. Participants who became more accepting of regulation as deliberation progressed were persuaded by the potential harms from lack of antibiotic oversight. Some of these participants supported a boundary midway between guidelines and regulation as it allows some freedom of choice; however, these participants struggled to provide a concrete example of this type of boundary. When discussing Part 2 of the case study, many participants were persuaded when they considered analogous situations of generally accepted regulations that benefit society (e.g., vaccinations, traffic lights, and speed limits). For example, one participant said,

“I change[d] my mind...I’m still very resistant to it.... [During the] last meeting I was just so bent on I want to have the freedom to put my foot down and say I want this antibiotic, but that 5,500 deaths of MRSA has just been rolling around in my head and maybe—it’s hard because I don’t like the idea of regulation and things like that. Although you’re right, the vaccinations and things, I was vaccinated for polio and I think everyone should,” (CD group in Silver Spring, MD).

3. Participants who firmly did not want regulations were concerned about: (1) the practicality and feasibility of regulations; (2) absolute regulations that were rigid or had unintended consequences—such as preventing those who need a treatment from getting it; and (3) external, profit-motivated, and/or authoritarian forces determining the regulations. Of note, some of these participants acknowledged potential benefits from regulating antibiotic use.

Participants who acknowledged the need for boundaries suggested increased “accountability” of doctors by verifying that a bacterial infection exists before prescribing antibiotics or tracking the number of antibiotic prescriptions. These participants believed accountability for doctors would reduce poor prescribing decisions. Some of these participants questioned “blind faith” in doctors and doubted that education would change doctor behavior.

Participants also discussed how penalties or incentives could increase accountability. Some participants suggested constructive improvement strategies (e.g., continuing education classes), whereas others called for stricter penalties (e.g., fines). Other participants said these penalties could hinder a physician's ability to use education and experience to make recommendations. For example, one participant said, "*The doctors are educated and they make their recommendations and they should be making them based on the information at hand.... If their hands just continually get tied with more and more and more regulations, they'll be less able to do that,*" (CD group in Durham, NC).

Case Study 3. Obesity Management: Comparing Treatment and Prevention

We used the *Obesity Management* case study only in the Citizens' Panel method at all four locations. Participants primarily deliberated about this case study during breakout groups with a facilitator in each group.

This case study provided background information about the obesity problem in the United States, health problems associated with obesity, current approaches for managing obesity (e.g., diet, exercise, and surgery), and current approaches for preventing obesity (e.g., changing school/community environments and taxing unhealthy foods). After introducing the case study and hearing from three expert witnesses (a bariatric surgeon, a public health expert, and a law professor who supported limited government involvement in medicine), we divided participants in North Carolina, Maryland, and California into three breakout groups to discuss the case.

In the case study scenario, participants could spend money from a federal grant to choose three of the following six options to reduce obesity: (1) intensive diet and exercise programs for the moderately obese, (2) bariatric surgery for the moderately obese, (3) change the community environment, (4) change the school environment, (5) increase taxes on non-nutritious foods for people who buy them, and (6) increase taxes on companies that make non-nutritious food.

In Illinois, the case study was structured differently: It did not include the six closed-ended options described above, but rather an open-ended question about the need to spend money from a federal grant. Further, facilitators left the room to allow the full group to deliberate about the case study together; notetakers remained in the room to capture discussion. However, the activity did not work as intended because participants became disorganized. Therefore, we revised the structure and organization of the case study for the North Carolina, Maryland, and California sessions.

Main Themes That Emerged During Deliberation

In this section, we summarize main themes that emerged from deliberation of this case study, organized by overall beliefs about obesity, followed by responses to the six options presented in the case study.

Participants' beliefs about obesity often shaped their understanding of this case study and the overall issue.

Most participants believed that obesity was a matter of choice determined by individual behavior. These participants explained that people caused their own obesity by making poor choices. As such, participants argued that people who are obese need to take personal responsibility for their choices and health care decisions. One participant explained, "Let's boil it down to personal responsibility. You don't need a million rules. You don't need government intervention....You don't need to tell them they're not going to have chicken.... Give this person the responsibility," (CP group in Chicago, IL).

The majority of participants believed that adults who were obese were unlikely to change their lifestyle; therefore, preventing children and young adults from becoming obese was the best way to handle the obesity epidemic. Participants believed that video games, unhealthy food, and a lack of understanding about the dangers of obesity caused the obesity problem.

These participants explained that although adults were obese resulting from a lack of personal responsibility, children become obese by living in a society that made it difficult to be active and to eat well.

Option1: Intensive Diet and Exercise Program. Despite evidence that these programs were less successful than surgery, participants were overwhelmingly in favor of this option. In every breakout group, the majority chose this option. Participants did not typically explain why diet and exercise were the best option, but they often believed that education about diet and exercise would encourage people to change their behaviors. Of note, participants often interpreted this option as applying to every individual, not just the moderately obese. Also, participants appeared to interpret this option as preventive measure (not just treatment). Participants frequently cited benefits to these programs that the case study and experts did not mention, such as improved overall health. A few participants were concerned about why the evidence said diet and exercise were generally ineffective when using weight loss as the outcome. These participants often questioned the evidence's accuracy and noted that weight loss was not the only or most important outcome. One participant explained that *"you don't have to necessarily lose weight to be healthier"* (CP group in Sacramento, CA).

Only a few participants cited issues with Option 1 and explained that people already know how to eat well and exercise. For example, one participant noted, *"I think there is nobody out there who doesn't know that eating vegetables and not eating processed foods and being active instead of being sedentary is the key to a healthy lifestyle...everyone looks at education as the key to it but I do think...we know how to live a healthy lifestyle.... Implementing that knowledge is, I think, where the disconnect lies"* (CP group in Silver Spring, MD).

Option 2: Bariatric Surgery for the Moderately Obese. Participants were almost unanimously opposed to using Federal grant money to fund surgery for the moderately obese. These participants explained that surgery would only affect a few people, whereas all of the other options would affect more individuals. Additionally, insurance would cover bariatric surgery, so it was unnecessary for the grant to cover the treatment. Further, participants noted that surgery complications (which the case study explained occur in 88 percent of individuals), were so harmful that surgery was less beneficial and more harmful than diet and exercise.

However, a few participants wanted to provide surgery for the moderately obese; these participants noted that due to genetics, obese individuals are not always at fault. Other participants were often skeptical of this argument and struggled to believe that some individuals could not lose weight if they truly tried diet and exercise options. Below is a deliberative exchange portraying this argument (CP group in Durham, NC):

Female 1: I may want to include [Option] 2. Because I can't relate. I don't have a gene in my DNA for obesity. But some people do. And no matter how much they push back from the table, no matter how much they don't drink soda and they eat crackers, they still can't lose weight.

Female 2: Do you honestly believe—

Female 1: Yes I do. There's one or two that goes to my church. And I'm not being funny. When he sits down it's three chairs.

Female 2: And you honestly believe that person nowhere at no time is cheating?...

Female 1: Listening to the different testimonies that he's given, I believe he's done all he can to lose weight and he just has not been able to.

Option 3: Changing the Community Environment. Participants overwhelmingly favored changing the community environment to provide people with more opportunities to be healthy. Participants favored changing the community environment because they believed that many people were unable to exercise due to a lack of safety in public parks and a lack of access to private gyms.

Option 4: Changing the School Environment. Participants were unanimously in favor of changing the school environment, frequently expressing that most eating and exercising habits were formed as children. Almost all participants wanted to implement more physical education courses, healthier lunch options, fewer vending machines with unhealthy food options, and more education about healthy eating in schools. In the California Citizens' Panel, the facilitator questioned the large group about why they were willing to restrict children's choices when they were reluctant to implement anything mandatory or overly coercive for adults. Participants did not find it problematic to restrict choices for children, citing the fact that parents already restrict children's choices or that these types of activities did not actually limit choice.

Options 5 and 6: Increase Taxes on Non-Nutritious Foods for People Who Buy Them and Increase Taxes on Companies that Produce Non-Nutritious Foods. A majority of participants were immediately opposed to any taxation and immediately discarded options 5 and 6. These two options are summarized together because most participants believed taxing companies that produce unhealthy food would lead to an increase in cost passed down in taxes to the consumer.

Most participants argued that taxes would not work, claiming that people would purchase what they wanted. Some participants cited previous interventions, such as mandatory calorie counts on soda bottles, as evidence that people would consume what they wanted despite government intervention. Other participants noted that taxation penalized even non-obese people who enjoyed unhealthy food in moderation. Some participants argued that it was unfair to take away the freedom to eat what people wanted, making claims such as *"This is America. It was founded on the ability to do as you like, to let the top down, bench seats, no seatbelts, 65 miles an hour, baby"* (CP group in Sacramento, CA). Additionally, some participants thought that using grant money to lobby for tax increases seemed like an inappropriate use of funds.

Despite this dominant viewpoint, a few participants wanted to consider taxation. Most of these participants made the analogy to cigarettes, claiming that the effect of taxes on cigarettes and tobacco companies has had a major impact on decreasing smoking. Some of these participants explained that they personally made poor eating choices because unhealthy food is cheap; if this food were more expensive, they would be less tempted to purchase it. Some of these participants also expressed that if government was paying for the price of the obesity epidemic, it had the right to intervene.

Case Study 4. Heart Disease Treatment: Comparing Medicines Only and Stents Plus Medicines

This section summarizes background and participant responses to the *Heart Disease* case study. We used this case in the Citizens' Panel method at all four locations. On Day 1 of Citizens' Panel, participants discussed background information in the large group and deliberated during smaller breakout groups. The full group later reconvened, and a representative from each breakout summarized their discussion.

The *Heart Disease* case study provided background information about using medicines only or stents plus medicine to treat mild heart disease; the case also explained that patients who had stents also had to take medicines indefinitely. Stents are a more costly option and have more risks, but offer faster relief. The case study also explained that the evidence suggests that 33 percent of people with mild heart disease should receive a stent as the first-line of treatment, but 55 percent of people currently receive stents as the first option. With this information, we asked participants to assume the role of a patient advisory board to a health insurance plan; participants then deliberated about what advice and recommendations they should provide to the health plan.

Main Themes That Emerged During Deliberation

The following main themes emerged from deliberation of this case study:

Many participants discussed why more stents were being used than the evidence suggested was appropriate. Participants frequently began deliberating by attempting to identify the responsible party for the overuse of stents. Some participants blamed doctors who wanted to make profits. Other participants blamed patients who wanted a “*quick fix*,” with one participant saying “*because it’s the American way*” to want immediate relief (CP group in Sacramento, CA). Some participants believed patients’ lack of information about the risks and benefits associated with stents led to unnecessary stent use. Participants cited both the physician and patient as responsible for gathering the necessary information to make informed decisions.

Participants’ perceptions of the risks associated with stent surgery appeared to influence their views on why patients would have the procedure. Participants who found the stent procedure too risky struggled to understand why patients would want a stent. These participants frequently voiced that if the doctor educated the patient about the risks, the patient would choose not to have one. One participant explained, “*I would hope that the facts would scare the person away from doing it*,” (CP group in Silver Spring, MD). Other participants believed that stents were not too risky and that patients would want a stent because benefits were more immediate. Some of these participants acknowledged that for patients with certain occupations (e.g., a firefighter, professional athlete, or factory worker), the benefits of stents might outweigh the risks.

Participants appeared willing to recommend some boundaries based on the evidence of overuse of stents as first-line treatment. Several participants preferred education and preservation of choice, but given the potential cost burden on others, many participants acknowledged the need for limits on patients and/or increased monitoring of doctors. Yet, participants struggled with the feasibility of these boundaries and how they would be implemented in the real world.

Information about the cost of stent procedures appeared to persuade most participants to recommend reducing the number of stents to more closely match what the evidence suggested. Most participants cited cost as their primary concern, including ramifications of certain people using more than their fair share of resources. One participant explained, *“I think what needs to be the concern is whether we need to do [something] collectively to try to get the cost down so that everybody can get the same health coverage,”* (CP group in Silver Spring, MD).

Although participants overwhelmingly recommended that insurance companies reduce the number of stents being placed, they differed in what steps should be taken to lower the number of stents. Participants suggested the following actions to curb stent use:

Implement stricter guidelines on who can receive stents. Most participants suggested stricter guidelines on who could receive stents. These participants often argued there should be set criteria (e.g., family history, medical history, and extent of blockage) for who could receive a stent as a first-line treatment. Some participants wanted to cede control over these criteria to experts.

Implement permeable boundaries/guidelines to reduce stent use. However, almost all participants wanted exceptions to these stricter guidelines. The vast majority of participants wanted a review board that would allow some individuals to receive a stent before the guidelines would allow; some participants wanted to give doctors the ability to circumvent the guidelines without needing prior approval if they provided justification.

Educate patients about risks and benefits. Participants who believed that patients would not want stents if they had more information frequently noted that educating patients would suffice to lower the number of stents without necessarily having guidelines.

Participants differed greatly on whether patients who had a difficult time making lifestyle changes were more or less deserving of a stent. Some participants believed that patients who did not make lifestyle changes and accept personal responsibility for their heart disease should not expect a stent. One participant explained, *“Smoking, obesity, your diet, those are lifestyle things, and if a patient’s unwilling to do that, I think you have to draw a line somewhere and say we can’t keep going on like this.... We’re talking about the costs here, it’s like whatever I want, whatever I need without regards to at least being accountable first...we’re a society, people have to pay eventually for whatever it is we want,”* (CP group in Silver Spring, MD). However, other participants disagreed and said those patients who were unable to make lifestyle changes could benefit the most from receiving a stent because they were unlikely to improve with only medication.

Case Study 5. Comparing Approaches to Preventing Illness: A Fictional Case

Only participants in the Citizens' Panel method discussed the *Fictional Case* during the last session on Day 2 of deliberation. We did not use this case in Chicago's Citizens' Panel due to time constraints; therefore, only participants in the Citizens' Panel sessions in North Carolina, California, and Maryland deliberated about this case study.

This case study asked participants to make treatment decisions regarding a fictional illness called Shake, Rattle, and Roll (SRR). In this case study, patients who contract SRR develop severe symptoms and become completely dependent on others for their personal care. Additionally, a person would have a 1 in 50 chance of developing this fictional illness. As part of deliberation, participants discussed three approaches to manage SRR:

- **Option 1** focused on treating this illness with an expensive medication (\$25,000 per patient per year) that only was successful in treating 50 percent of the cases.
- **Option 2** focused on prevention by encouraging everyone to exercise 30 minutes per day, every day. In this option, public health officials estimated that only 50 percent of the population would exercise every day and that it would cost \$1,500 per person to train the public to do the exercises. As such, this was the most expensive option.
- **Option 3** focused on prevention by putting an inexpensive medicine in the water supply (\$5 per person). However, 1 out of every 1,000 people who drank the water would eventually die from an allergic reaction. This was the cheapest option.

During deliberations, a facilitator stood next to flipcharts showing brief descriptions of each option. Participants rotated through each flipchart to discuss the merits and drawbacks of each option with any participants who were standing near the same flip chart. Then, in the large group, participants discussed which option they would chose and why.

Main Themes That Emerged During Deliberation

The following main themes emerged from deliberation of this case study:

Many participants struggled to understand the facts in the case study. For example, many participants thought choosing Option 2 (exercise and treatment) meant individuals who were unable or unwilling to exercise would immediately contract SRR. Instead, that person would have a 1 in 50 chance of contracting SRR.

Participants also asked for and had questions about details that were not in the case study. For example, participants were unclear on what would happen if they did not drink the water every day, what constituted exercise, how one caught the illness, and if SRR was contagious.

The majority of participants preferred Option 3 (water supply) because it had the lowest costs and fewest people affected by SRR. Participants preferred this option despite the 1 in 1,000 chance of death from an allergic reaction and cited several reasons for this choice.

Weighing quality of life versus length of life. Participants weighed the quality of life versus the length of life in making their decision. Participants differed on whether they would prefer to live with a debilitating illness or to die. More participants indicated they would prefer to die than live an unpleasant life; for example, one participant said, *"I'd rather be dead, [than]*

to be living [like that, affecting the people of my family,” (CP group in Durham, NC). The participants who wanted to live frequently cited hope for a cure as a key motivation for their preference.

Assessing the number of people affected. Most participants cited the dramatically higher number of people that SRR affected because of the exercise/treatment option to justify choosing the preventive option. For these participants, many people having poor quality of life was worse than a few people dying.

Identifying practical concerns about exercise. Some participants expressed concerns that everyone might not be able to exercise (as in Option 2). Participants did not clearly articulate whether they thought these individuals would consequently contract SRR, but overall participants did not want to have an accident and suddenly be susceptible to a disease.

Considering costs. A few participants mentioned that the higher costs of Option 2 (treatment and exercise) were a significant barrier. When the facilitator prompted participants to explore if opinions would change if this option were cheaper, few participants changed their mind.

Considering freedom of choice. Participants who chose Option 2 (water supply) rarely changed their mind even when others pointed out that this decision restricted choice. Some participants noted that generally, individuals must take personal responsibility for their health; however, many participants who chose Option 3 said the exercise and treatment option put those who were unable to exercise at an unfair disadvantage. Participants also noted that choice would be restricted regardless of which option they chose, but did not elaborate further.

A minority of participants chose Option 2 (exercise and treatment) because it preserved life and they questioned the safety of medicine in the water supply as in Option 3.

Preserving life. Many participants supported Option 2 because no one died. Some of these participants were concerned that every child born would have a 1 in 1,000 chance of dying from an allergic reaction in Option 3 (water supply).

Safety of medicine in the water supply. A few participants expressed concern about the long-term safety of medicine in the water supply and found it hard to believe the evidence citing its safety.